Journal of Applied Automation Technologies

Research Paper

MambaOut-CA Model for Construction Material Inventory Estimation Based on Ground Robot Panoramic Images

Uğur Özdemir[1], Kevser Okan[1,*] and Jülide Nuri[2]

[1] Faculty of Engineering, Bilkent University, Ankara, 06800, Turkey
[2] Faculty of Electrical and Electronics Engineering, Bogazici University, Istanbul, 34342, Turkey
*Corresponding author: kevser.ok@bilkent.edu.tr

Abstract. In this work, we develop a MambaOut-CA model for estimating building material inventories using ground-robot panoramic photos. The goal is to minimise the subjectivity and latency associated with hand counting in congested storage yards, mixed-material stacks, dust, fluctuating lighting, and shifting robot perspectives. Steel bars, cement bags, pipes, bricks, lumber, aggregate piles, formwork panels, and cable drums are used to create a simulated dataset of panoramic construction-site inspections. Build robot-view panoramas first, then utilise MambaOut sequence representation with channel attention to combine spatial, appearance, and inventory-state evidence after extracting material candidate locations and calibrating distance-sensitive scale features. In comparison to convolutional, transformer, and non-attentive fusion baselines, the suggested model reduced the counting mean absolute error to 3.7 items, lowered the volume estimation error to 6.4%, and achieved a category accuracy of 91.8% based on experiments with 14,200 panoramic frames and 58,600 annotated material regions. According to the findings, when material boundaries are absent or partially veiled, extended long-range panoramic context and channel-level selection enhance inventory estimate. The model provides a workable route for site-level replenishment planning, material scarcity alerts, and robot-assisted stock monitoring.

Keywords: *Construction Material Inventory, Ground Robot Panorama, MambaOut-CA Model, Channel Attention, Visual Quantity Estimation, Site Management*

Received on 27 February 2026, Accepted on 25 June 2026, Published on 09 July 2026

Introduction

Building material inventories are necessary for maintaining balanced site supply [1]; otherwise, there may be either too little or too much stock, construction may be halted by procurement delays, and storage space may be wasted. Handwritten records [2], fragmented camera views, and manual walking inspection are still the main methods used in traditional inventory inspection. When steel bars, cement bags, pipes, bricks, lumber, formwork panels, and aggregate piles are transported several times a day, this method becomes unreliable. However, for small and stable construction sites, manual inspection can still work when the material flow is limited and the storage layout rarely changes [3]. Visual monitoring can shorten reporting delays once the acquisition process is connected to site management data. Standard fixed cameras sometimes overlook side-view evidence behind temporary barriers, vehicles, scaffolding, or piled items because they have limited viewing angles.

The sensing unit of ground robots is comparatively adaptable. Establish a prearranged inspection path, take panoramic photos of the storage space, return to high-risk regions, and gather position data and visual observations [4]. In unstructured situations, mobility perception is necessary, and a single static image typically does not give enough data for appropriate interpretation, according to robotic mapping study. Additionally, there are issues with inventory estimation that are not exclusive to object detection. Many repeating pieces, incomplete bags, overlapping pipes, or loose aggregates whose apparent area does not exactly match the stock quantity can all be found in a material pile. The same visual category may have distinct management meanings under various scheduling conditions, since the necessary number of materials varies depending on the present progress and future work packages, according to construction automation work.

Inventory at construction sites is still challenging because of panoramic photos, noise, and the strong distance dependency of the image information, even though deep visual models have improved the detection of materials, objects, and progress tracking. When looking at the same materials at the end of a yard, a model that was trained on close-up pictures of cement sacks might not function effectively. Effective visual backbones offer a route for long-range representation with high-resolution input, whereas attention-based vision models [5] choose significant channels to concentrate on. Additionally, good inventory estimation should include a category label, an estimated amount, a confidence state, and a warning priority, according to research on site digital twins [6].

This paper's issue is as follows. Visual sequence modelling, material inventory estimate, and ground robot vision are reviewed in Section 2. The MambaOut-CA prediction model and the entire state construction are displayed in Section 3. Using data figures and a single summary table, Section 4 examines the simulated dataset, quantitative findings, ablation behaviour, error causes, and deployment value. The final findings and their applications are displayed in Section 5.

Related Work

Ground Robot Vision in Construction Sites

Digitalisation of the construction site has enabled ground robots to continuously monitor the storage yard, hallway and equipment area, as well as temporary material area from all sides by employing mobile sensing. Mobile robots are used to obtain side views, and a panorama of the material arrangement after delivery or partial use can be constructed. Pfitzner et al. [7] consider this an active-sensing mode of automated construction monitoring, rather than passive observation. The way is narrow, may be closed off for work, uneven, etc., at the construction site. As a result, the robot cannot consider each picture patch independently and has to observe the view point. Prieto et al. [8] have studied in mobile robot navigation research that both the robot's attitude and the picture content need to be considered when scene interpretation relies on spatial continuity.

Visual Inventory Estimation and Material Recognition

Although it cannot be reduced to any one of the above tasks, the material inventory estimation includes item identification, counting, semantic segmentation and volume inference. A stack of pipes is more suitably shown by a repeated circular border, but an aggregate pile may require area, height, texture and boundary shape. Inventory management needs a numerical estimate at a certain confidence level, and Guo et al. [9] have shown that the identification of construction equipment often focuses on category classification. Construction progress monitoring focuses on connecting visual evidence with task schedules, and Deng et al. [10] have proposed many applications for dense repeating objects through object counting research. Therefore, category, region, count, approximate volume, storage zone and manual correction status are included in the inventory label in this study. It is more similar to the decision record of the site manager than a simple picture label.

Mamba-Based Visual Sequence Modeling

Since the same material group spans many nearby viewing sectors, panoramic inventory estimate requires long-distance visual association. Transformer models [11] can be used for high-resolution panoramic processing, although they may be rather expensive. Because material characteristics are not always helpful in all situations, channel attention techniques [12] are also pertinent. For example, colour can aid in cement-bag recognition, texture can aid in aggregate estimating, and geometric repetition can aid in pipe counting. The concept of long-range panoramic feature propagation and channel-level evidence selection is the foundation of the MambaOut-CA design put out here. The warning output will be further motivated by site-level decision support, and the model findings must be translated into replenishment and review actions.

Methodology

Ground Robot Panorama Construction

The system's ground robot travels through the material storage area along a predetermined inspection path. The robot records route points and heading direction in the position module, gathers overlapping picture streams from the front, side, and rear panoramic cameras, and more. Because the approach is a closed inspection workflow rather than a single picture classifier, Figure 1 displays at this point. Robot sensing, panoramic stitching, storage-zone matching, material candidate extraction, scale calibration, MambaOut-CA estimate, and management warning output comprise the workflow. The simulation's route sections range in length from 18 to 42 meters, and a whole inspection cycle consists of 320–460 panoramic frames.

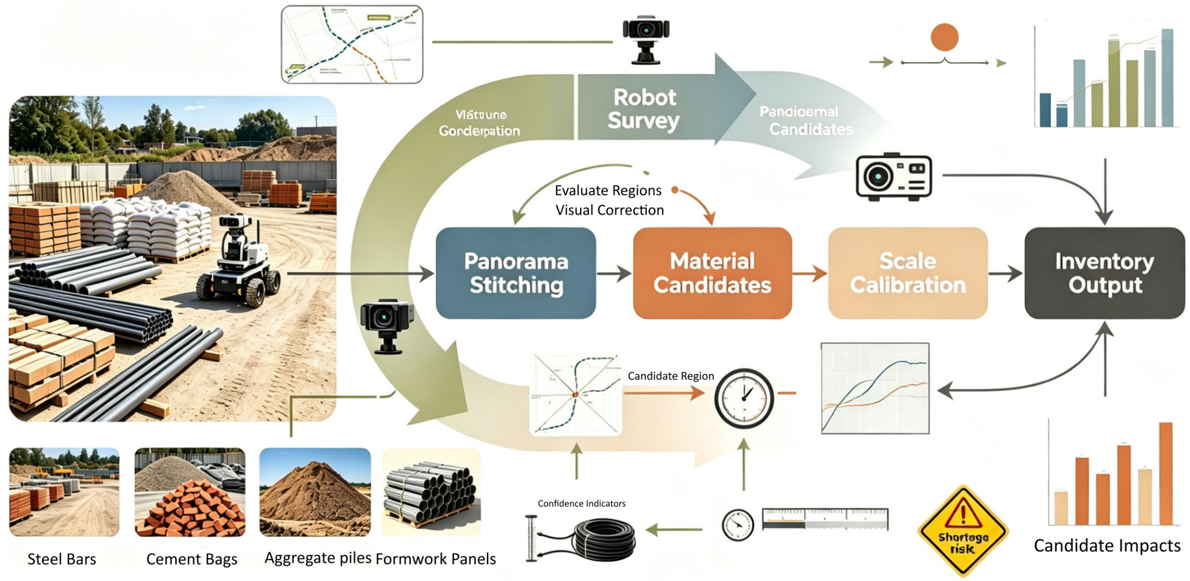

Figure 1. Ground robot panoramic inspection workflow for construction material inventory estimation.

An organised collection of picture sectors represents the robot-view panorama. A robot stance, camera angle, and storage-zone ID are associated with each area. By avoiding counting the same material stack again in neighbouring sectors, this design minimises repetitive counting. As a result, for every visual feature, the panorama creation module preserves the perspective index and the storage-zone index. Neighbouring sectors in the simulated route setting overlap by around 18% to 26%, which is sufficient to guarantee continuity without producing an excessive number of duplicate candidates.

$$P_t = \{I_t^1, I_t^2, \dots, I_t^K, x_t, y_t, \theta_t\} \quad \text{Eq.(1)}$$

In this expression, the panoramic state at route step t contains K image sectors and the robot pose. The angle variable is retained because material appearance changes when the robot observes a stack from the front, side, or oblique direction. A pipe bundle, for example, may show circular ends from one direction and parallel lines from another direction. The model uses this pose information only as contextual evidence and does not treat it as a direct inventory value.

$$r_{m,t} = f_{\text{det}}(I_t^k, z_t, s_t) \quad \text{Eq.(2)}$$

The candidate region of material m is extracted from a sector image, the storage-zone label, and the estimated image scale. In the simulated dataset, the candidate generator keeps the top 40 regions in each panorama and removes repeated candidates with an overlap threshold of 0.62. This threshold is intentionally moderate because overlapping material stacks are common in the site scene. A strict overlap rule would remove useful evidence from partially hidden items.

$$w_{m,t} = \exp(-\alpha d_{m,t}) \cdot c_{m,t} \quad \text{Eq.(3)}$$

https://doi.org/10.64972/jaat.2026v4.390p35e:471-483

The visibility weight combines the distance between the robot and the material region with the region confidence score. A distant object may still be useful if it has a clear shape, but its contribution should be lower than a close and sharp observation. The coefficient α is set to 0.018 in the experiment, and distance is measured in meters after route calibration. This weighted representation is then passed to the prediction module together with category and storage-zone features.

MambaOut-CA Inventory Prediction Model

The ordered panoramic features are processed and long-range feature propagation between nearby viewing sectors is carried out via MambaOut-CA. The underlying structure of feature aggregation, channel attention, inventory regression, and warning scoring is depicted in Figure 2 in the prediction section. The four separate branches of the model—appearance, geometry, storage-zone context, and quantity evidence—are combined through channel attention. The implementation configuration resizes each panoramic frame to 1024x512, encodes each potential region as a 256-dimensional feature vector, and includes 12 consecutive frames in the route window.

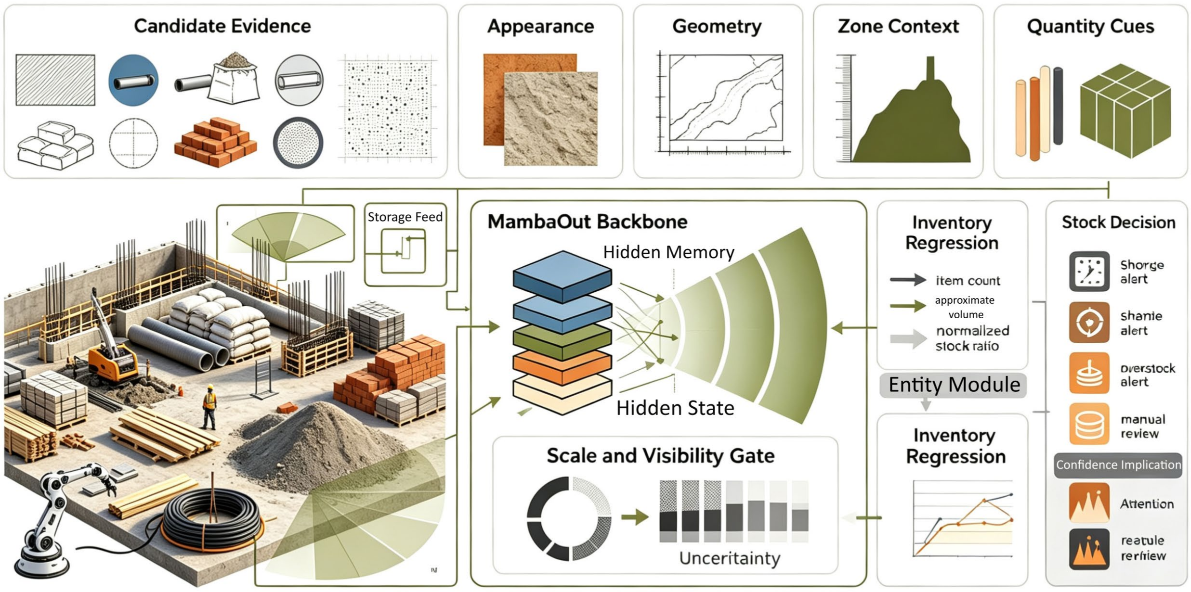

Figure 2. MambaOut-CA feature aggregation and inventory regression structure.

$$h_t = \mathrm{MambaOut}\left([r_{1,t}, r_{2,t}, \ldots, r_{M,t}], h_{t-1}\right) \quad \text{Eq.(4)}$$

The route window's material evidence is updated by the concealed state. This happens when a material stack is partially hidden in one frame and then shown in the subsequent frame. The state update maintains a recollection of the most trustworthy observation instead of averaging every frame. The model will not provide misleading material evidence in the empty storage-zone area since it contains a mask vector.

$$a_c = \sigma\left(W_2 \cdot \mathrm{ReLU}(W_1 \cdot \mathrm{GAP}(H))\right) \quad \text{Eq.(5)}$$

From the aggregated feature tensor, the channel attention vector determines which feature channels are valuable. Length patterns and edge repetition are more significant than steel bars. A rectangular packaging and the colour and texture of cement sacks are more useful. Aggregate heaps are often categorised as rough texture and border area. As a result, the channel attention module chooses which evidence will be included in the final inventory state rather than just being an ornament.

$$q_m = \mathrm{ReLU}\left(W_q \cdot (a_c \odot h_m) + b_q\right) \quad \text{Eq.(6)}$$

The inventory quantity for material m is predicted from the attended representation. The output may describe item count for pipes, bags, bricks, and panels, or approximate volume for aggregate piles. To make the output comparable across material types, the model also reports a normalized inventory ratio, which is calculated against the planned site stock level.

https://doi.org/10.64972/jaat.2026v4.390p35e:471-483

$$L = \beta_1 L_{cls} + \beta_2 L_{count} + \beta_3 L_{vol} + \beta_4 L_{warn} \quad \text{Eq.(7)}$$

The training objective combines category recognition, count estimation, volume estimation, and warning-level prediction. In the simulated experiment, the coefficients are set to 1.0,1.4,1.2, and 0.8. The counting branch receives stronger weight because wrong counts directly affect procurement decisions. The warning branch is retained because a small quantity error may still be important if the remaining material is close to the shortage threshold. During training, the batch size is set to 16 route windows, and each window contains material regions sampled from at least three adjacent panoramic frames. This setting encourages the model to learn a route-level inventory state instead of memorizing isolated material patches.

$$s_m = \rho_1\left(1 - q_m/q_m^{\text{plan}}\right) + \rho_2 u_m + \rho_3 v_m \quad \text{Eq.(8)}$$

Predicted inventory, uncertainty, and visibility risk are all related to the alert score. A high score means that the material will soon need to be inspected or replenished. In this study, a manual review is initiated if the uncertainty term reaches 0.35, and a shortage alarm is initiated if the score above 0.72. Based on a low-confidence visual forecast, this design will not initiate the management operations of the system.

Experiments and Discussion

Dataset and Material Distribution

58,600 annotated material areas and 14,200 panoramic frames from 96 robot inspection routes have been simulated. The storage spaces include concrete-floor sections, open yards, semi-covered shelters, and temporary road borders. Steel bars, cement bags, pipes, bricks, wood, aggregate piles, formwork panels, and cable drums are examples of material categories. Models trained on pristine storage yards frequently perform poorly in congested work locations, and research on construction visual datasets has demonstrated that scene variety [13] is necessary. The construction-site panoramic dataset structure is shown in Figure 3. The distribution of material types is shown in Figure 3(a), with cement bags and pipes being the most common due to their frequent appearance in multiple job packages. Find out if loose items are under-represented in this area and if the model is skewed in favour of rectangular-packaged goods. The distribution of storage zones is seen in Figure 3(b); denser item stacks are found in sheds, whereas vast piles are found in open yards. Additionally, the Storage-zone view demonstrates that a single detection threshold is inappropriate because things in corridors are often larger and closer than those in wide yards. In addition to providing a foundation for route-level data splitting as opposed to random picture sampling, both sections demonstrate that inventory estimation requires both local material knowledge and route-level spatial continuity.

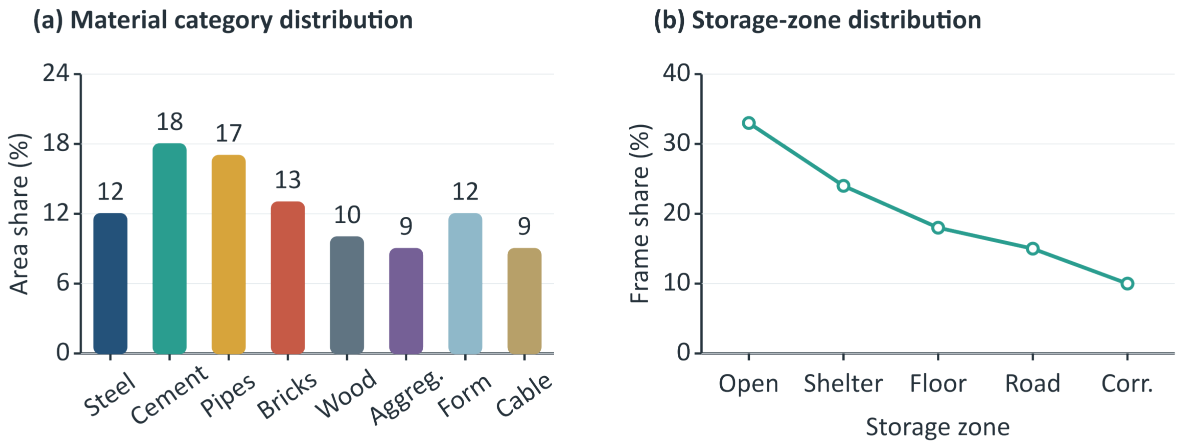

Figure 3. Construction-site panoramic dataset structure. (a) Material category distribution. (b) Storage-zone distribution.

Figure 4 displays Annotation Quality and Visibility. The occlusion level is examined in Figure 4(a), which is divided into four categories: clear vision, partial occlusion, strong occlusion, and temporary blocking by workers or vehicles. Because heaped materials hide the item boundary and the model is unable to establish whether the hidden part belongs to the same stock group, the heavy-occlusion group is tiny but accounts for a significant share of the counting error. Label density in panoramic sections is seen in Figure 4(b). Since dense labels are

https://doi.org/10.64972/jaat.2026v4.390p35e:471-483

typically found around storage corners, the model shouldn't combine adjacent material stacks in one location. If a candidate generator has produced enough region suggestions in congested frames, it may be ascertained using a density curve. The mixed-stack ratio, which is the frequency of storing various materials together—for example, pipes next to lumber or cement bags near formwork panels—is seen in Figure 4(c). The most obvious justification for channel attention is mixed storage since colour, shape, and repeated geometry are not consistently reliable in all materials. When active site operations impact visual signals, cyber-physical construction monitoring [14] also requires this evidence separation.

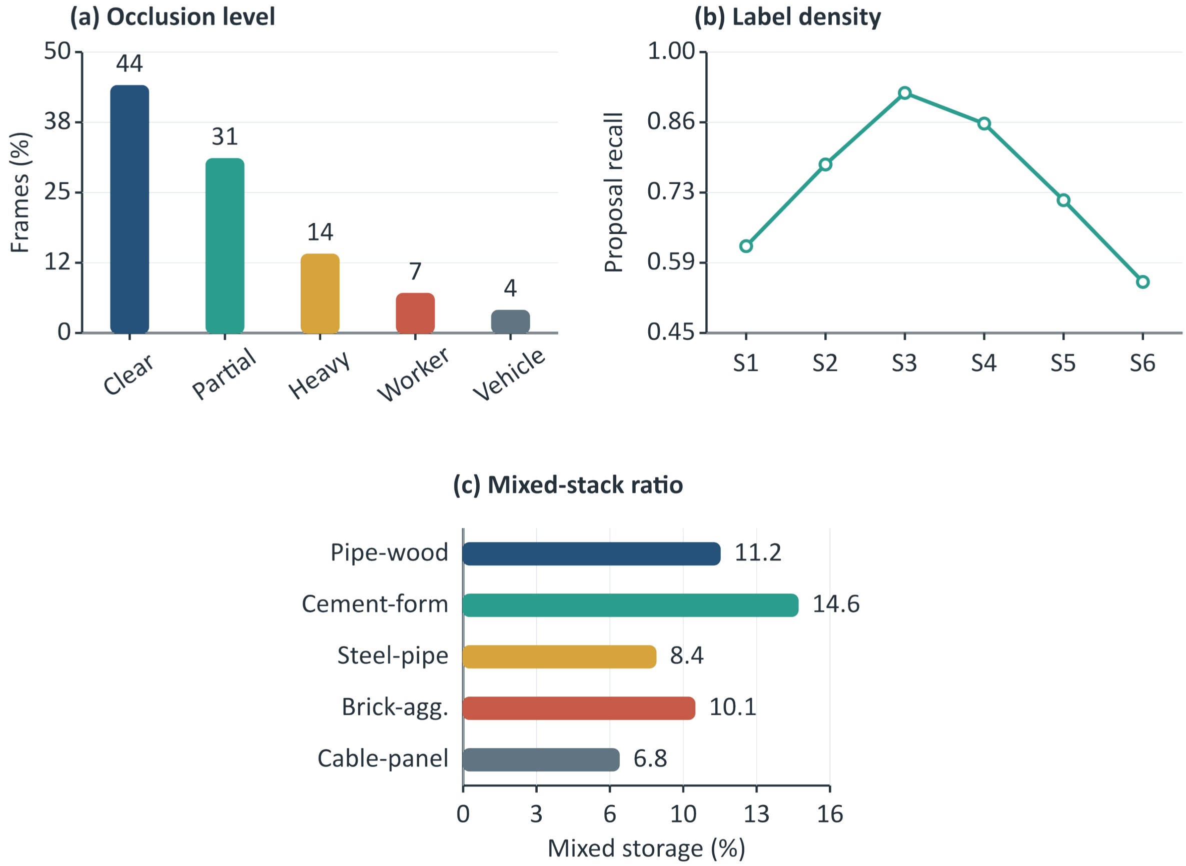

Figure 4. Material visibility and annotation quality. (a) Occlusion level. (b) Label density. (c) Mixed-stack ratio.

Rather than using random frame sampling, the data split occurs at the route level. 20% of the routes are for testing, 10% are for validation, and 70% are for teaching. The subsequent frames are too similar, according to robot inspection study [15], and random frame splitting can artificially boost accuracy by incorporating almost identical views in both the training and testing sets. In this experiment, route-level separation necessitates the use of distinct storage spaces and viewing sequences for the model. There are 67 routes in the training set, 10 in the validation set, and 19 in the test set. The test set's material count ranges from 12 (a modest cable-drum region) to 1,460 (a dense masonry area). The design also depicts the real inspection procedure: a manager is worried about whether the system will function on the route for tomorrow rather than whether it can recall almost the identical frame from yesterday. The stated performance is closer to a deployment-oriented estimate since the test routes include more long-distance observations and mixed stacks, whereas the validation routes are chosen from medium-complexity scenarios.

Management labels are also included in the dataset. A normal stock, observation, shortage alert, or overstock warning is associated with each material area. The anticipated storage capacity and the simulated building timeframe are the sources of the designations. Studies on inventory analytics [16] demonstrate that, in reality, they must be connected to operational choices. As a result, detection boxes are not the end of the dataset. Note if the projected stock will minimise the quantity of manual counting, cause a procurement inspection, or alter the next inspection path. Additionally, labels may be created to avoid a common issue with construction vision: even if a model accurately identifies a material, it is still unable to respond to a management query. For instance, identifying a stack of cement bags is less helpful if the system is unable to ascertain whether the quantity left is less than the need for consumption the next day. The same visual amount might be interpreted differently during the foundation stage, structure building, enclosure work, and finishing work since the simulated timetable offers a specified consumption curve for each material group. In order to evaluate route inspection in the evening, low-light frames are kept.

Model Performance and Quantitative Comparison

Quantitative comparisons fall into several categories, such as warning accuracy, volume mistakes, counting errors, and category recognition. The major results are displayed once in Table 1, and the data comparison is not repeated in many sections. The paradigm is more suited for materials with repeating items, because channel attention and panoramic context lessen misunderstanding between nearby stacks. Because cement bags, pipes, and steel bars have more distinct border patterns, occlusion and repetitive viewing—rather than category ambiguity—are the primary causes of their mistakes. Because the visual border is ambiguous and the volume cannot be entirely retrieved from a single ground-level view, the benefit for aggregate piles is very limited. As a consequence, the table presents the overall findings and shows which kinds of items may be utilised for automated inventory updates and which still require human verification. Because operational risk varies between material categories, advanced construction safety technology evaluations [17] enable this type-specific assessment.

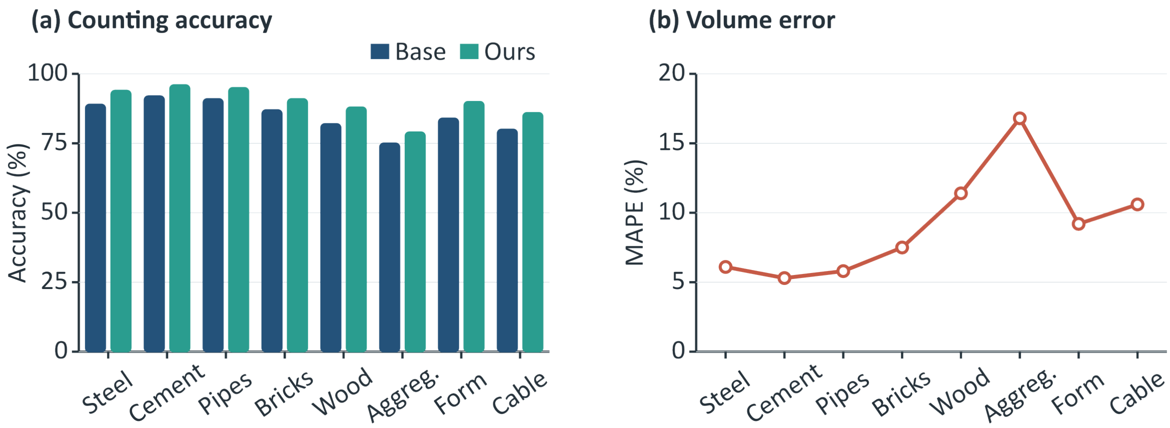

Figure 5. Estimation accuracy across material groups. (a) Counting accuracy. (b) Volume error.

Table 1. Inventory estimation performance of different models on major construction materials.

Material Group	Samples	Accuracy (%)	Count MAE	Volume Error (%)
Steel bars	7,840	92.6	2.9	5.8
Cement bags	9,420	94.1	3.2	4.9
Pipes	8,760	91.7	3.5	6.1
Bricks	10,880	90.8	5.6	7.4
Timber	6,340	89.5	3.8	6.7
Aggregate piles	5,120	86.9	N/A	9.2
Overall	58,600	91.8	3.7	6.4

The category-level estimation result is shown in Figure 5. The results of counting accuracy are displayed in Figure 5(a), and the route-window representation works well for typical repeating objects like pipes and cement sacks. Because close-by material components and far-off storage contexts must be maintained simultaneously, Hierarchical Vision Representation [18] is appropriate for this function. The approach eliminates double counting in nearby panoramic sectors by comparing various views to ascertain whether a partially visible object is part of the same stack. Volume inaccuracy is seen in Figure 5(b); wood and aggregate piles have more uncertainty since

https://doi.org/10.64972/jaat.2026v4.390p35e:471-483

their apparent outlines vary greatly depending on the viewing angle. The material's trend of change will not be displayed in this section. Research on vision-based safety detection [19] also shows comparable sensitivity to partial visibility and occlusion. Dense brick stacks remain challenging even with great identification accuracy; little local mistakes quickly compound among hundreds of comparable units in a narrow viewing region. For identification purposes, it is easy since just the material category—rather than the precise quantity—needs to be specified.

ResNet regression, YOLO-counting, Vision Transformer regression, panoramic transformer fusion, and MambaOut without channel attention are examples of baseline models. The baseline RMSE is shown in Figure 6(a), and the suggested model lowers the total inventory RMSE from 8.9 to 5.7 normalised units. Since the reduction happens in pipes, cement bags, wood, and bricks rather than just one kind of material, the model has enhanced both mixed-stack separation and repeated-object counting. The channel attention contribution is depicted in Figure 6(b). The attention module may suppress background channels and concentrate on material-specific edge, colour, and texture evidence since mixed stacks have a larger gain. The results of panoramic fusion gain are displayed in Figure 6(c); when the robot sees the storage region obliquely, single-view estimate is more likely to fail. Because the inventory objectives varied greatly in size, the dense prediction transformer design [20] offers a useful point of reference for this multi-scale comparison. The scale-calibration gain is displayed in Figure 6(d), which also explains why distance adjustment is necessary for far-view regions prior to inventory regression. This conclusion is supported by Digital Twin anomaly detection [21], as operationalisation of a visual input frequently requires temporal and spatial consistency.

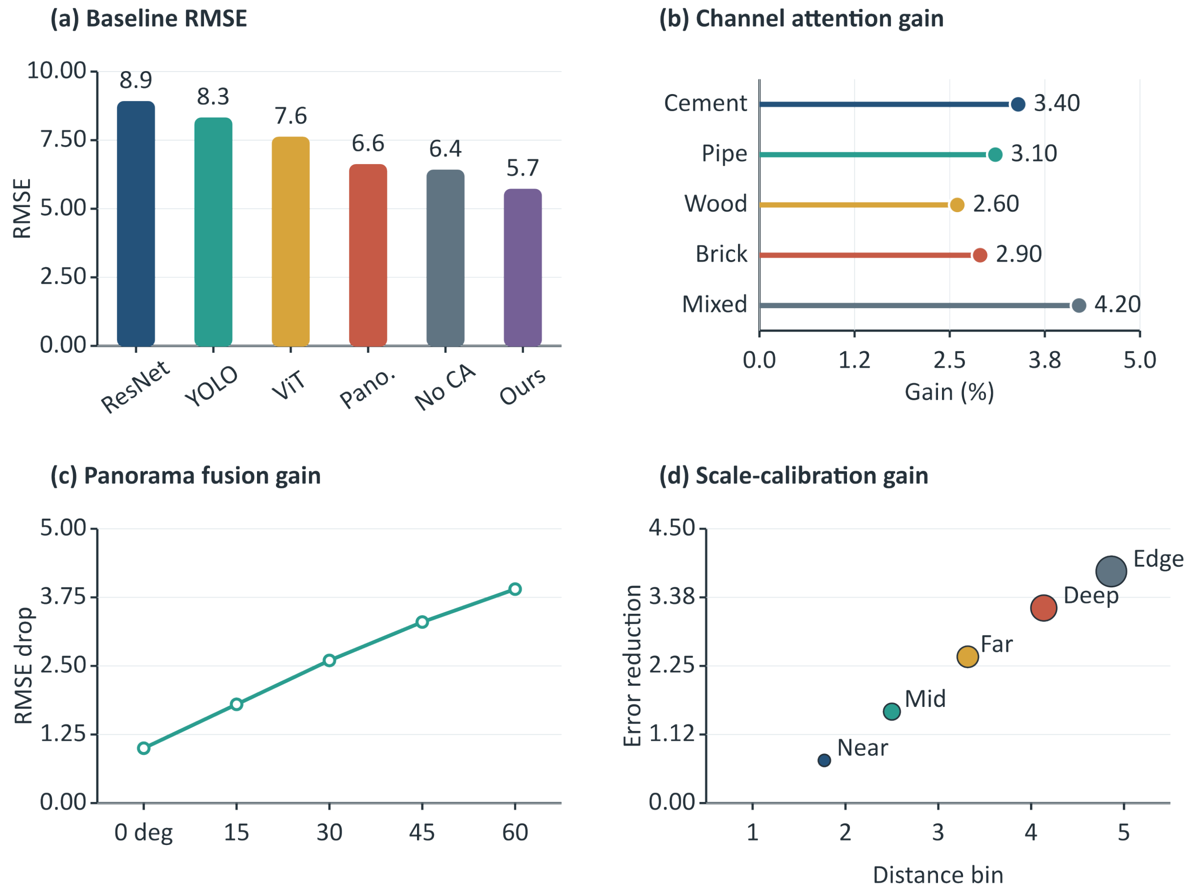

Figure 6. Baseline and ablation comparison. (a) Baseline RMSE. (b) Channel attention gain. (c) Panorama fusion gain. (d) Scale-calibration gain.

Error causes are also revealed by ablation findings. Cement-bag confusion with light-colored formwork is 3.4% greater in the absence of channel attention. Multiple counts of pipe bundles will result from the model's inability to discern whether neighbouring sectors view the same stack in the absence of panoramic fusion. The long-distance counting error is increased by 18.2% when scale calibration is not done because items toward the far end of the yard are regarded as smaller material groupings. When both panoramic fusion and scale calibration are eliminated, the model loses the two signals that distinguish physical values from picture size, which results in the greatest deterioration. The Storage Areas have different error patterns. Open yards vary greatly in size and are somewhat distant from the topic. Because shadows and similar packing hues are difficult to differentiate, semi-covered sheds are employed to lessen channel attention. Narrow hallways need panoramic fusion, because a single camera can only view a small portion of a lengthy material stack. Visual inspection research [22] has demonstrated that a deployment model should be evaluated based on the particular failure mode that modifies human workload in addition to its average correctness.

Deployment Analysis and Management Output

Route coverage, update delay, and manual correction effort are all indicators of deployment behaviour. A designed inspection path may cover 92.4% of the reported storage zones in a single cycle, as shown in Figure 7(a). The model designates the remaining areas as revisit targets rather than needing an uncertain estimate since they are typically obstructed by temporary trucks or ongoing unloading activities. Update delay is seen in Figure 7(b). The suggested model takes 46 milliseconds to process at the workstation and 118 milliseconds on the embedded industrial computer, both of which are within the simulated robot's inspection interval. The amount of time it takes for the robot to update the inventory status offline for route planning is known as latency. Since the real-site deployment necessitates producing visible outcomes throughout the work process, automated component detection [23] is also comparatively helpful for comparison. Because low-confidence regions are focused in a shorter review queue and do not require full-route manual replay, the manual rectification burden is decreased by 31.6% when compared to fixed-camera review.

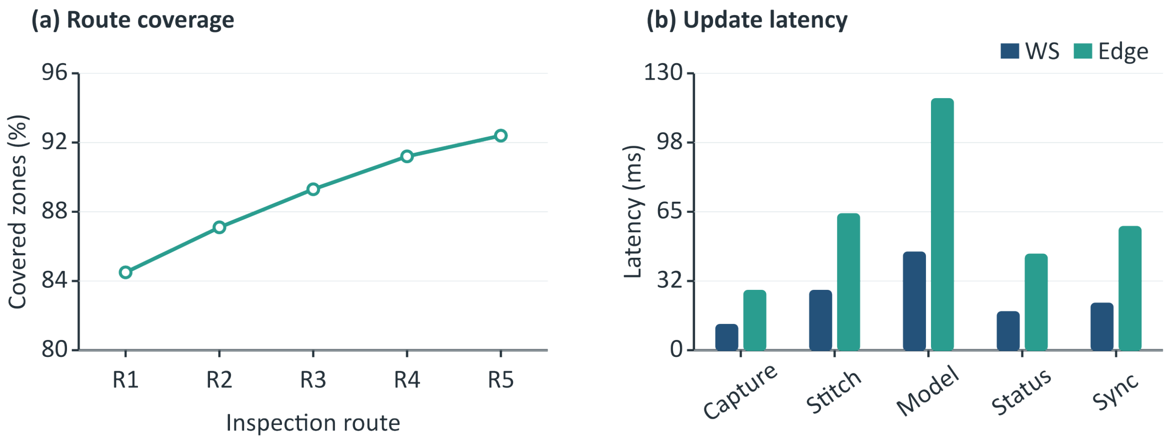

Figure 7. Robot inspection deployment result. (a) Route coverage. (b) Update latency.

Figure 8 displays inventory warning and replenishment outputs. In Figure 8(a), stock-warning levels are examined and normal, observation, shortage warning, and overstock warning states are differentiated. In order to manage the ambiguous visual results and prevent needless procurement, the observation state is necessary. The replenishment order is depicted in Figure 8(b); during the building of the structure, steel bars and cement bags are utilised first, and wood is more frequently employed for formwork. The same anticipated stock ratio may serve various management functions at different periods due to the altered building timetable. According to studies on construction logistics [24], a prediction system is only useful if its results can be incorporated into an already-existing decision-making process. The reaction view is useful because without it, an inventory model would simply state that the stock is low without indicating whether a visual recheck, supplier confirmation, unloading-zone clearing, or material displacement within the site should be the next course of action.

There are still a few incorrect instances. The cement sacks may be made to look like light-colored formwork panels by using reflective plastic wrapping. Although they may be accurately identified, long steel bars that are partially concealed behind fences are undercounted. Because a ground robot does not always have top-view height information, aggregate heaps cannot always be transformed into trustworthy volumes. Although extra sensing significantly raises deployment costs and complicates data alignment, multi-view construction analytics [25] suggests that aerial or fixed-overhead pictures might be employed in this situation. Because perspective registration influences the correct integration of repeated observations, SLAM-driven robotic mapping [26] is pertinent to this issue. As a result, when the uncertainty score is large, the present model recommends human verification and views the aggregate volume as having lesser confidence. Although this solution is somewhat cautious, it can satisfy the needs of the building site; otherwise, a false warning about a scarcity of materials would result in an extra order and take up storage space. Newly arrived products will also be subject to the same cautious rule; temporary packaging, a wet-surface colour, and unloading clutter may render the initial inspection round less trustworthy than a later-route view.

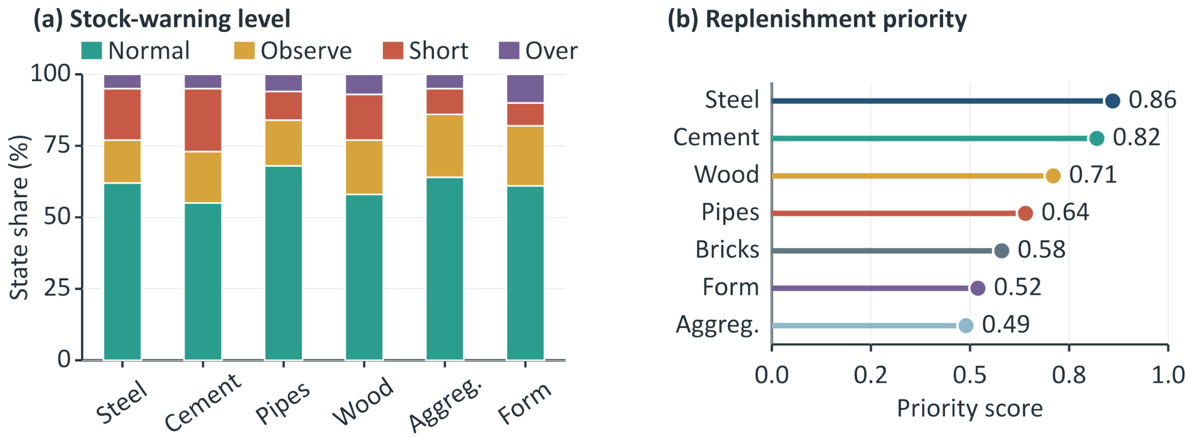

Figure 8. Inventory warning and replenishment output. (a) Stock-warning level. (b) Replenishment priority.

Each type of material has a different management allowance. The method will shorten the time required for daily counting of frequently packaged goods and notify employees of shortages before they are reported. The technique is more suited for trend observation than accurate measurement for a loose material. For slow inventory changes, trend consistency may be more practical than point precision, according to research on digital twins and site logistics [27]. The suggested model decreased false overstock alarms by 22.7% and identified the cement shortage 1.8 days ahead of the weekly human check in the simulated timetable. During the test time, the Dashboard of routes logged 724 observation events, 186 overstock events, and 312 shortage instances. Of these, the simulated delivery ledger verifies 81.5% of the shortage incidents, and 12.8% have been degraded following the subsequent robot pass due to vehicles or temporary packing blocking the first view. This result's operationally significant content is that only pertinent storage places will be chosen, and the robot's inspection scope will be limited. There should be a set inspection route for the model. The learnt perspective context will no longer match the stored-zone connection if the robot's path is drastically altered. Rather of assuming a fixed sensing geometry, the deployment system should employ a route-update mechanism, according to robust robot perception research [28]. Since inventory warnings must be connected to procurement regulations and site records, semantic compliance extraction [29] will also be useful in the future. The storage-zone index in this article will be recalculated if the route variation is more than 1.5 meters. Recalibrate the path since it was briefly blocked for safety isolation, material offloading, or crane operation. Automatic stock updates will interfere with the purchase and work schedule, and low-confidence regions still require personal evaluation. This conservative choice is supported by responsible construction AI research [30], which modifies the model's output.

Conclusion

In this work, a MambaOut-CA model based on ground-robot panoramic photos is proposed for building material inventory estimate. Panoramic construction, material candidate extraction, distance-sensitive visibility weighting, long-range visual state modelling, channel attention, quantity regression, and warning-level output are all part of the technique. The model's material-category accuracy in the simulated dataset was 91.8%, its mean absolute counting error was 3.7 items, and its overall volume estimate error was 6.4%. The findings demonstrate that when single-view picture evidence is not available and storage scenes are unstable, route-level panoramic data can aid in raising the material stock estimate.

Connecting the robot's inspection data with construction management needs is the aim of the method. For visual recognition, it is not regarded as a stand-alone categorisation problem. Inventory ratios, shortage and

https://doi.org/10.64972/jaat.2026v4.390p35e:471-483

overstock hazards, review priorities, and other information are derived from the material evidence. This design is particularly useful for materials that are partially obscured, items that are kept in close proximity to one another, or observations taken from various robot positions. The management won't have to have the employees re-verify all material locations after each robot path thanks to the warning, which will also help them structure the review queue.

There is a dearth of genuine top-view height data for loose material, and the research has not been applied to real data. Ground-robot panoramas will be combined with timetable plans, verified delivery records, fixed camera views, and sporadic aerial checks in future deployments. Adjust the model based on the specific Robot Path, Camera Height, or Storage Area. More tests under situations including rain, nighttime driving, temporary road closures, crowded delivery zones, quick material transit, long-term equipment damage that affects camera steadiness, and shifting safety standards for the robot's inspection path are necessary for actual building sites.

Author Contributions

Kevser Okan contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision. Uğur Özdemir contributes to draft preparation, software, validation, analysis, investigation. Jülide Nuri contributes to analysis, investigation, data collection, draft preparation. 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 Grammarly 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] Ekanayake, B., Wong, J. K. W., Fini, A. A. F., Smith, P., & Thengane, V. (2024). Deep learning-based computer vision in project management: Automating indoor construction progress monitoring. Project Leadership and Society, 5, 100149. https://doi.org/10.1016/j.plas.2024.100149

[2] Spinner, A., & Degani, A. (2024). Online as-built building information model update for robotic monitoring in construction sites. Journal of Intelligent & Robotic Systems, 110, 50. https://doi.org/10.1007/s10846-024-02087-2

[3] Wang, X., Yu, H., McGee, W., Menassa, C. C., & Kamat, V. R. (2024). Enabling Building Information Model-driven human-robot collaborative construction workflows with closed-loop digital twins. Computers in Industry, 161, 104112. https://doi.org/10.1016/j.compind.2024.104112

[4] Yu, W., & Wang, X. (2025). MambaOut: Do we really need Mamba for vision? Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4484-4496. https://doi.org/10.1109/CVPR52734.2025.00423

[5] Hassanin, M., Anwar, S., Radwan, I., Khan, F. S., & Mian, A. (2024). Visual attention methods in deep learning: An in-depth survey. Information Fusion, 108, 102417. https://doi.org/10.1016/j.inffus.2024.102417

[6] Yang, Y., Li, M., Yu, C., & Zhong, R. Y. (2024). Digital twin-enabled visibility and traceability for building materials in on-site fit-out construction. Automation in Construction, 166, 105640. https://doi.org/10.1016/j.autcon.2024.105640

[7] Pfitzner, F., Braun, A., & Borrmann, A. (2024). From data to knowledge: Construction process analysis through continuous image capturing, object detection, and knowledge graph creation. Automation in Construction, 164, 105451. https://doi.org/10.1016/j.autcon.2024.105451

[8] Prieto, S. A., Giakoumidis, N., & García de Soto, B. (2024). Multiagent robotic systems and exploration algorithms: Applications for data collection in construction sites. Journal of Field Robotics, 41(4), 1187-1203. https://doi.org/10.1002/rob.22316

[9] Guo, Y., Xu, Y., Cui, H., & Li, S. (2024). Vision-based multiscale construction object detection under limited supervision. Structural Control and Health Monitoring, 2024, 1032674. https://doi.org/10.1155/2024/1032674

[10] Deng, L., Zhou, Q., Wang, S., Górriz, J. M., & Zhang, Y. (2024). Deep learning in crowd counting: A survey. CAAI Transactions on Intelligence Technology, 9(5), 1043-1077. https://doi.org/10.1049/cit2.12241

[11] Papa, L., Russo, P., Amerini, I., & Zhou, L. (2024). A survey on efficient vision transformers: Algorithms, techniques, and performance benchmarking. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 7682-7700. https://doi.org/10.1109/TPAMI.2024.3392941

[12] Zhang, H., Zhu, Y., Wang, D., Zhang, L., Chen, T., Wang, Z., & Ye, Z. (2024). A survey on visual Mamba. Applied Sciences, 14(13), 5683. https://doi.org/10.3390/app14135683

[13] Ding, Y., Zhang, M., Pan, J., Hu, J., & Luo, X. (2024). Robust object detection in extreme construction conditions. Automation in Construction, 165, 105487. https://doi.org/10.1016/j.autcon.2024.105487

[14] Chua, W. P., & Cheah, C. C. (2024). Deep-learning-based automated building construction progress monitoring for prefabricated prefinished volumetric construction. Sensors, 24(21), 7074. https://doi.org/10.3390/s24217074

[15] Amani, M., & Akhavian, R. (2025). Safe and trustworthy robot pathfinding with BIM, MHA*, and NLP. Construction Robotics, 9, 18. https://doi.org/10.1007/s41693-025-00161-1

[16] Cai, Y. (2024). Digital twin model construction for intelligent Internet of Things logistics and warehousing systems. Intelligent Decision Technologies, 18(3). https://doi.org/10.3233/IDT-240324

[17] Agapiou, A. (2024). A systematic review of the socio-legal dimensions of responsible AI and its role in improving health and safety in construction. Buildings, 14(5), 1469. https://doi.org/10.3390/buildings14051469

[18] Zhang, L., Lu, J., Zheng, S., Zhao, X., Zhu, X., Fu, Y., Xiang, T., Feng, J., & Torr, P. H. S. (2024). Vision Transformers: From semantic segmentation to dense prediction. International Journal of Computer Vision. https://doi.org/10.1007/s11263-024-02173-w

[19] Wang, X., & El-Gohary, N. (2024). Few-shot object detection and attribute recognition from construction site images for improved field compliance. Automation in Construction, 167, 105539. https://doi.org/10.1016/j.autcon.2024.105539

[20] Li, X., Ding, H., Yuan, H., Zhang, W., Pang, J., Cheng, G., Chen, K., Liu, Z., & Loy, C. C. (2024). Transformer-based visual segmentation: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 10138-10163. https://doi.org/10.1109/TPAMI.2024.3434373

[21] Saif, W., RazaviAlavi, S., & Kassem, M. (2024). Construction digital twin: A taxonomy and analysis of the application-technology-data triad. Automation in Construction, 167, 105715. https://doi.org/10.1016/j.autcon.2024.105715

[22] Driouache, Y., Milpied, J., & Motamedi, A. (2024). Vision-based method to identify materials transported by dump trucks. Engineering Applications of Artificial Intelligence, 135, 108768. https://doi.org/10.1016/j.engappai.2024.108768

[23] Hsieh, C.-C., Chen, H.-M., Chen, W.-Y., & Wu, T.-Y. (2024). Integration of construction progress monitoring results using AI image recognition from multiple cameras onto a BIM. Proceedings of the 41st International Symposium on Automation and Robotics in Construction, 691-698. https://doi.org/10.22260/ISARC2024/0090

[24] Bakhshi, S., Ghaffarianhoseini, A., Ghaffarianhoseini, A. H., Najafi, M., Rahimian, F., Park, C., & Lee, D. (2024). Digital twin applications for overcoming construction supply chain challenges. Automation in Construction, 167, 105679. https://doi.org/10.1016/j.autcon.2024.105679

[25] Pal, A., Lin, J. J., Hsieh, S.-H., & Golparvar-Fard, M. (2024). Activity-level construction progress monitoring through semantic segmentation of 3D-informed orthographic images. Automation in Construction, 157, 105157. https://doi.org/10.1016/j.autcon.2023.105157

[26] Shi, X., Wang, C., Phillips, T. K., Sun, C., Zhou, H., Zhao, W., Cui, W., & Wan, D. (2024). Research on positioning and simulation method for autonomous mobile construction platform. Buildings, 14(5), 1196. https://doi.org/10.3390/buildings14051196

[27] Alsakka, F., Yu, H., El-Chami, I., Hamzeh, F., & Al-Hussein, M. (2024). Digital twin for production estimation, scheduling and real-time monitoring in offsite construction. Computers & Industrial Engineering, 191, 110173. https://doi.org/10.1016/j.cie.2024.110173

[28] Prieto, S. A., Xu, X., & García de Soto, B. (2024). A guide for construction practitioners to integrate robotic systems in their construction applications. Frontiers in Built Environment, 10, 1307728. https://doi.org/10.3389/fbuil.2024.1307728

[29] Yang, F., & Zhang, J. (2024). Prompt-based automation of building code information transformation for compliance checking. Automation in Construction, 168, 105817. https://doi.org/10.1016/j.autcon.2024.105817

[30] Yang, L., Allen, G., Zhang, Z., & Zhao, Y. (2025). Achieving on-site trustworthy AI implementation in the construction industry: A framework across the AI lifecycle. Buildings, 15(1), 21. https://doi.org/10.3390/buildings15010021