Joint CSCE Construction Specialty & CRC Conference 2025
Conférence conjointe spécialisée en construction de la SCGC et CRC-2025

Montreal, Quebec
July 28-31, 2025 / *28-31 juillet 2025*

Improving machine utilization in off-site construction using computer vision methods

Xue Chen[1]*, Weining Zeng[1], Runcong Liu[1], Mohamed Al-Hussein[1], and Ahmed Bouferguenou[2]

[1] Department of Civil and Environmental Engineering, University of Alberta, Edmonton AB, T6G 2R3, Canada
[2] Campus Saint-Jean, University of Alberta, Edmonton AB, T6C 4G9, Canada

ABSTRACT: On construction sites, heavy machinery such as mobile cranes, large cutting machines, and forklifts are often shared among multiple workstations due to their high costs. The efficient allocation of these shared machines significantly impacts financial investment, machine utilization, and overall construction productivity. If machine capacity is insufficient to meet the demands of multiple workstations, non-productive delays may arise due to waiting times. Conversely, an excessive number of shared machines can lead to resource inefficiencies and unnecessary investment costs. Therefore, evaluating and optimizing machine utilization is crucial for improving existing off-site construction factories and designing new ones. This study develops a vision-based framework to analyze machine utilization from two key perspectives: machine utilization rate and machine waiting time. The optimal number of machines and workstations is determined by maximizing utilization without introducing additional waiting time across multiple workstations. To achieve this, machine operation states are automatically identified using a machine mapping method based on hybrid object-tracking results. Machine demands at different workstations are predicted through an integrated analysis of object detection results and standard work procedures. By comparing machine operation states with workstation demands, machine waiting time and utilization rates are computed. A simulation model with various machine-workstation configurations is then designed to enhance utilization and minimize waiting times. Finally, the proposed vision-based evaluation and optimization model is validated through a practical off-site construction case study.

1. INTRODUCTION

In off-site construction factories, building components are manufactured in parallel in different workstations. Large machines, such as overhead cranes, forklifts, nailing and cutting bridge, are usually shared by multiple workstations in the factory to save financial investment and workspace. For instance, an overhead crane used to lift completed building components can be used by floor panel workstations and roof workstations together. A cutting machine may work in multiple adjacent wall construction stations in parallel. In this working scenario, the workstations used for constructing difference building components are fixed, while the shared machines can move among multiple workstations either along the pre-installed tracks (as in the case of overhead crane and nailing/cutting bridge) or across designated pathways (in the case of forklifts and transport trucks). This sharing mechanism introduces a delicate balance. On the one hand, when the shared machines' work capacities are much higher than the demands of workstations, machines will be in an idle state frequently, leading to wastes in the utilization of work resources. For example, an overhead crane often sits unused because component lifting is a relatively infrequent task, resulting in wasted investment and underutilized equipment (overproduction). On the other hand, waste due to waiting for machines may occur if the required machines cannot be acquired in time due to their busy schedule of jobs at multiple workstations. If a machine is tied up servicing another workstation, it may lead to non-productive waste, such as worker idle time, delayed production steps, and potential disruption of the assembly schedule in the current workstation. To be specific, if multiple wall stations simultaneously require the cutting machine, some workstations may be forced to pause operations, leading to bottlenecks and increased cycle times. Therefore, the appropriate configuration of the shared machines with the corresponding workstations and the allocation of the machines among these workstations would have a great influence in the construction efficiency and the utilization rate of work resources.

To evaluate the performance of the shared large machines across multiple workstations, this study conducted a comparison between the machines' work states and the workstations' demands based on actual construction data. Traditionally, construction information is collected by manual observation and records, which is labor- and time-consuming for bulk data acquisition and analysis. With the application of surveillance cameras and computer vision models, this study works on automatic data collection from off-site construction factories and makes a data-driven analysis of machine utilization rate and waiting time for machines. A detection-based machine tracking model and a detection-based workstation requirements prediction model are developed using the YOLO V8 object detection model. Outputs from the computer vision-based models include machine work conditions, such as real-time locations of machines and their operation information (idle or in operation), and multiple workstations' work states, such as real-time work activities of each workstation and the required resources, which assist project managers get remote and automatic control of machine and workstation utilization conditions. Enabled by the data collected with computer vision methods, a simulation model covering the interested machines and workstations is further developed to optimize machine utilization rate and reduce productivity loss. Average task durations and requirements for machines per task (that are predefined within the Standard Operating Procedure) are inputted to the simulation model, in which different numbers of workstations and shared machines are tested. Based on the simulation results, the optimum configuration of shared machines and workstations is selected to improve the machine utilization rate and eliminate the waiting time for machines.

Overall, three objectives are achieved in this study: 1) developing a computer vision-based framework for construction data collection and machine requirement prediction, 2) conducting machine utilization rate evaluation and waiting time analysis across multiple workstations, and 3) improving work productivity with optimized configuration of machines and workstations in off-site construction factories.

2. LITERATURE REVIEW

Object tracking technology in the construction field uses a variety of model frameworks to adapt to different application scenarios. The detection-based tracking (TBD) framework mainly relies on detection algorithms, YOLO and Faster R-CNN and combines DeepSORT and Kalman Filter for object tracking. It is suitable for sparse object monitoring and short-term occlusion processing, such as real-time worker positioning and drone inspections. Joint detection and tracking (JDT) frameworks, such as FairMOT and JDE, use an end-to-end approach to improve tracking efficiency and are suitable for dynamic tracking of construction equipment and multi-target dense scenes. Transformer-based tracking models, such as TransTrack and TransMOT, use a spatiotemporal attention mechanism to enhance long-term tracking capabilities and are suitable for worker behavior analysis and multi-device interactive tracking in complex environments. Behavior recognition and pose estimation models such as Two-Stream CNN, ST-GCN, and OpenPose are suitable for construction action classification and safety compliance detection, enhancing the understanding of worker behavior. Multi-modal fusion models combine vision with LiDAR, RFID, BIM, and other technologies to improve the robustness of target tracking in harsh environments, which are suitable for night monitoring, high-precision positioning, and construction progress visualization (Guo et al. 2022; Li et al. 2024a). The current commonly used recognition algorithm in research is YOLO. In recent years, research applying the YOLO algorithm has mainly focused on scenarios such as construction and prefabrication plants. The TiDA system based on YOLOv4 focuses on collecting production time for prefabricated buildings. It triggers timing by detecting targets such as workers and equipment, with an average absolute error of between 0.72 and 0.96 minutes, which can meet the needs of production process management (Alsakka et al. 2023). In terms of construction activity recognition, YOLOv5 combines a re-identification algorithm to achieve an overall accuracy of 88.5% by detecting workers' posture and head direction, maintaining high recognition performance in complex construction environments (Li et al. 2024b). For safety compliance detection on construction sites, YOLOv8 achieves an mAP of 94.4% for recognizing personal protective equipment, effectively improving real-time safety supervision (Dimaano and Alon 2023). For multi-object detection and segmentation tasks, YOLOv11-Seg's mAP@0.5 is 0.808, and the F1 Score is as high as 0.8382, demonstrating good robustness in detecting sizeable mechanical equipment and nighttime monitoring (He et al. 2024). These studies still face challenges regarding occlusion, lighting changes, and data diversity. In response to these challenges, other studies have approached the problem from different perspectives. They use wearable sensors and existing low-resolution factory monitoring systems to automatically identify and track construction worker activities and prefabricated production progress. The deep learning-based method for identifying construction worker activities uses an IMU sensor to collect worker motion data and a neural network such as ConvLSTM for identification. Since this method does not rely on visual information, it can effectively avoid the challenges of insufficient lighting, occlusions, and restricted camera views, and achieve real-time monitoring of complex construction activities. Its 98.64% activity recognition accuracy shows the high feasibility and accuracy of deep learning in multi-sensor fusion (Karatas and Budak 2024).

High-cost heavy equipment is often shared among multiple workstations in construction and prefabrication plants. Fluctuating demand and fragmented scheduling can lead to notably low utilization levels, while queuing delays may account for a substantial portion of the cycle time (Song and Eldin 2012). In modular construction, non-value-added activities can represent a considerable share of total project costs, underscoring the importance of improving resource allocation and reducing idle time (Innella et al. 2019). In the absence of effective scheduling and coordination, even state-of-the-art equipment can remain underutilized or become significant bottlenecks (Arashpour et al. 2014). Consequently, understanding the factors influencing machine utilization is vital for devising robust optimization strategies. Despite the existence of various frameworks aimed at enhancing scheduling, heavy machine utilization typically hinges on a network of interrelated factors. Inefficient facility layouts or complex logistics can also impede equipment movement—forklifts in tight aisles or cranes located far from storage areas lead to extended travel times and increased switching penalties (Jiang et al. 2022). Unforeseen breakdowns or maintenance, especially for machines approaching their operational limits and requiring frequent inspections, further disrupt planning and reduce effective availability (Li et al. 2020). By identifying and quantifying these conditions, project teams can better anticipate inefficiencies and introduce more flexible scheduling mechanisms. Given these interconnected challenges, researchers have investigated a range of approaches to align equipment availability with workstation requirements. Linear and integer programming techniques, for instance, are commonly employed to allocate machines under relatively static conditions (Amiri et al. 2022). Discrete Event Simulation (DES) software captures the dynamic interplay among equipment, personnel, and layout constraints (AbouRizk et al. 2016). Iterative simulations help expose bottlenecks and enable the safe testing of alternative schedules. In precast operations, for example, DES-based scheduling has been shown to lower waiting times by reorganizing crane routes and re-sequencing production tasks (Araya 2022).

By structuring the literature review, this study found several research gaps: (1) Most current approaches focus on identifying resources (e.g., machinery, labor) rather than allocating them based on live machine statuses and concurrent workstation demands. As a result, there is a pressing need for an adaptive framework that draws on real-time tracking data, automatically evaluates multiple machine-workstation configurations, and applies dynamic optimization to simultaneously maximize utilization and minimize waiting times. (2) Traditional simulation methods, such as DES and mixed-integer programming, rely on historical data and fixed rules, making them less responsive to unmodeled real-time events. (3) It’s necessary to explore how to build an integrated framework combining computer vision analyzing results and simulation models to optimally allocates shared heavy equipment across multiple workstations, thereby improving large, shared machine operation condition.

3. METHODOLOGY

As displayed in Figure 1, the methodology developed in this study includes three main parts: a) a computer vision-based model for data collection, b) data-driven analysis on machine utilization rate and waiting time for machine, and c) simulation-based machine-workstation configuration optimization. In part a), the object detection model YOLO V8 is used to automatically identify the interested machines in multiple workstations, with processing speed taken into consideration. The model's performance is evaluated using a labeled dataset developed in the author's previous research (Chen et al. 2024), reaching a precision of 99.7% for nailing and cutting machine detection and 95.1% for hoisting crane detection. By connecting cameras installed at each workstation with the pre-trained YOLO V8 model, labels of interested machines and image coordinates from each video frame are automatically captured and stored in dataset. The image coordinates of machines are then converted to real-world coordinates using the homography matrix calculated in Equation [1] to map the image coordinate system to the real-world coordinate system.

$$[1]\ s\begin{bmatrix}x_{real}\\ y_{real}\\ 1\end{bmatrix} = \underbrace{\begin{bmatrix}H_{11} & H_{12} & H_{13}\\ H_{21} & H_{22} & H_{23}\\ H_{31} & H_{32} & H_{33}\end{bmatrix}}_{H}\begin{bmatrix}x_{image}\\ y_{image}\\ 1\end{bmatrix}$$

where s is the scalar used for scaling, H is the homography matrix, (x_{image},y_{image}) is image coordinate, and (x_{real},y_{real}) is real-life coordinate.

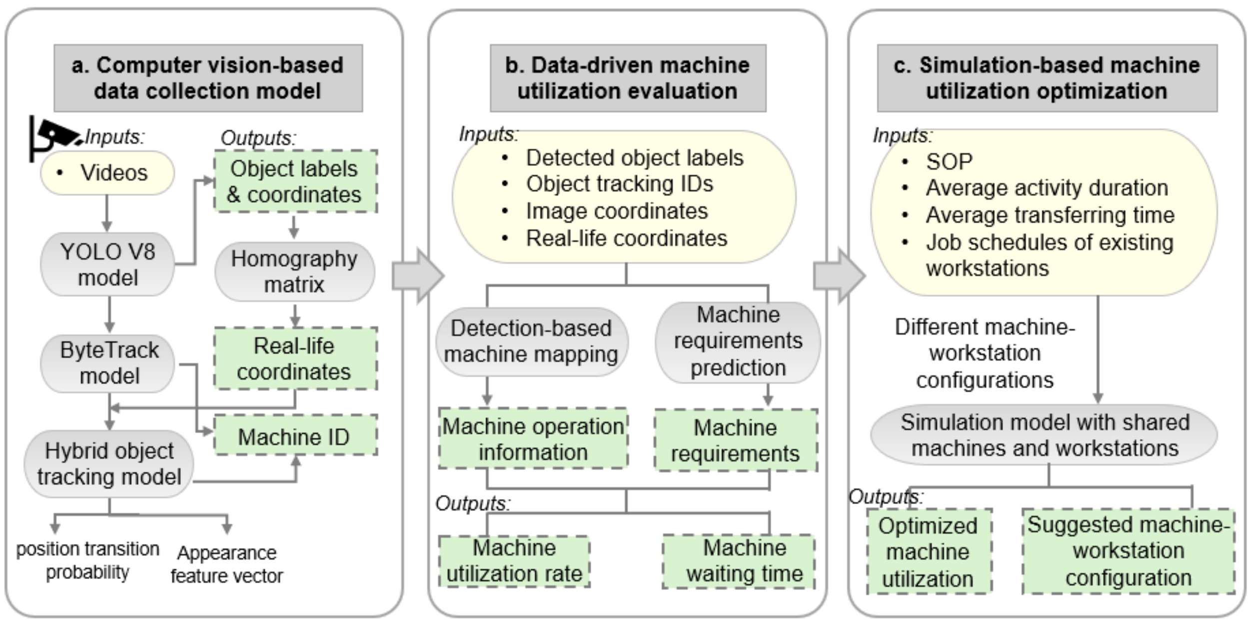

Figure 1: Overview of methodology

Employing the detection-based tracking method ByteTrack built in the YOLO V8 model, multiple interested machines are tracked under the same single camera by matching their positions across frames. However, how to match the detected machines and their positions across multiple cameras is still a challenge especially when the cameras don't have overlapping views. To solve this problem, a hybrid object tracking model combined with the graph-based position prediction method and the re-identification (Re-ID) method is developed in this study to assign the same global ID to the machine moving in multiple camera views. Firstly, a camera view map is designed for the graph-based position prediction method to distinguish the edge of each camera's field of view in the off-site construction site. When a machine moves near the edge of the current camera view, a position transition probability for its next camera view is given based on historical observations. Meanwhile, the feature vector of the machine appearing in the new camera view is extracted with the Re-ID model Omni-Scale Network (OSNet) to identify its appearance features. Finally, the similarity value is computed using the sum of position transition probability and the feature vector to assign object IDs, integrating appearance characteristics with spatial and temporal information in the hybrid object tracking model. The position transition probability only applies to objects that were previously detected near the camera's field of view edge within a defined time window.

To analyze machine operation conditions automatically, a tracking-based machine mapping method is then developed by integrating detected real-life coordinates with the construction site layout as listed in Equation [2]. For example, if the coordinate of crane stays within the scope of a floor workstation during a period, it will be predicted that the crane was busy with the tasks in the floor workstation at that time. Similarly, this rule also applies to other workstations. On the other side, if the interested machine keeps still at non-working areas, an idle state will be tagged to the machine. This position-judging process is achieved with Heron's formula in Python. Using the detection-based machine position mapping method, the operation information of the shared machines, such as idle, transferring, and busy, is tracked over time series.

$$[2]\ State_{machine} = \begin{cases} Idle \leftarrow if\ (x_{real}, y_{real}) \notin A_{working}\ and\ (x_{real\_t}, y_{real\_t}) - (x_{real_{t0}}, y_{real_{t0}}) \le \varepsilon, t \in [t_0, t_i] \\ Busy \leftarrow if\ (x_{real}, y_{real}) \in A_{working} \\ Transferring \leftarrow if\ (x_{real}, y_{real}) \notin A_{working}\ and\ (x_{real\_t}, y_{real\_t}) - (x_{real_{t0}}, y_{real_{t0}}) > \varepsilon \end{cases}$$

where $A_{working}$ is the working scope of multiple workstations, (x_{real_t}, y_{real_t}) are machine coordinates from t_0 to t_i, and ε is the threshold for judging where an object keeps still during a certain period.

Furthermore, machine requirements used for identifying potential waiting time for machines are predicted based on computer vision analyzing results and standard operation procedure (SOP). With the YOLO V8 object detection model, several unique work resources are automatically detected for the recognition of activities. For instance, material rack with joists is only used during the activity "joist placement". Once the material rack is detected from video frames, joist placement is assumed to be happening. This study only focuses on the activities completed with the assistance of machines and assumes that each workstation would manufacture building components following the predefined work sequences in SOP. Machine requirements are forecasted by identifying the start time points of the corresponding activities or the end time points of the preceding activities. For example, cutting and nailing machine is used in the activity "wall frame and plywood connection" at a wall construction station. The time point requiring this machine at the wall workstation is predicted by identifying the endpoint of its preceding activity "plywood placement". Based on real-time object detection results, when the tool (vacuum lifter) used for plywood placement disappears from the detection outputs, it's thought that the activity "plywood placement" has been finished. This time is then stamped as the timepoint when the nailing and cutting machine should be obtained to initiate the activity "wall frame and plywood connection". If this machine is not acquired at this time, waiting time for nailing and cutting machine will occur.

To validate the performance of the developed computer vision-based framework, surveillance videos, that serve as data inputs to the developed data-driven analysis model, are continuously collected from off-site construction sites. By applying the computer vision-based methods introduced above to the collected videos, machine operation information (including working states, working positions, and tracking IDs) and machine requirements of multiple workstations are automatically collected in time sequence from daily working processes. Machine waiting time of different workstations are recognized by comparing machine operation information with machine requirements of each workstation (see in Equation 3). Machine utilization rate is computed with Equation 4. The higher the machine utilization rate is, the better the machine is used in the off-site construction factory. However, too high a utilization rate may lead to more waiting time for machines at various workstations. Productivity loss will be further caused by waiting time for machine due to unreasonable configuration or schedule of shared machines and workstations. Therefore, this study evaluates machine utilization from two perspectives, machine utilization rate and machine waiting time, aiming to achieve machine utilization rate optimization without adding waiting time.

$$[3]\ Waiting\ for\ machine = \begin{cases} True \leftarrow if\ States\_Machine_i \neq Requirements\_Machine_i \\ False \leftarrow if\ States_{Machine_i} = Requirements\_Machine_i \end{cases}$$

where $States\_Machine_i$ represents machine working state and position at some point, and $Requirements\_Machine_i$ refers to workstation requirements for machine at the same time.

[4] $Machine\ utilization\ rate = \left. t_{busy} + t_{transferring} \middle/ t_{working} \right.$

where t_{busy} is the time that a machine is working at a workstation, $t_{transferring}$ is the time when a machine is moving in the construction site, and $t_{working}$ is the total working time. t_{busy} and $t_{transferring}$ are collected from computer vision results with Equation [2].

Finally, a simulation model is developed to identify the optimal configuration of machines and workstations on construction sites, offering guidelines for both new off-site construction factory setups and the optimization of existing sites. The model is designed based on the SOP provided by an off-site construction factory. Key inputs include average activity duration, machine transfer time, and machine requirements for each activity defined within SOP. By mirroring the construction workflow at each workstation, the simulation model tests various configurations of workstations and machines to evaluate machine utilization performance in a virtual environment. Taking a floor workstation as an example, the workflow within this workstation is shown in Figure 3. Lower simulation outputs—specifically, low machine utilization rates and short machine waiting times—indicate a greater potential for adding new workstations without overloading existing resources. Conversely, if machine utilization and waiting time are high, the results suggest that additional shared machines should be introduced to the simulation model in order to accommodate the production demands of multiple workstations.

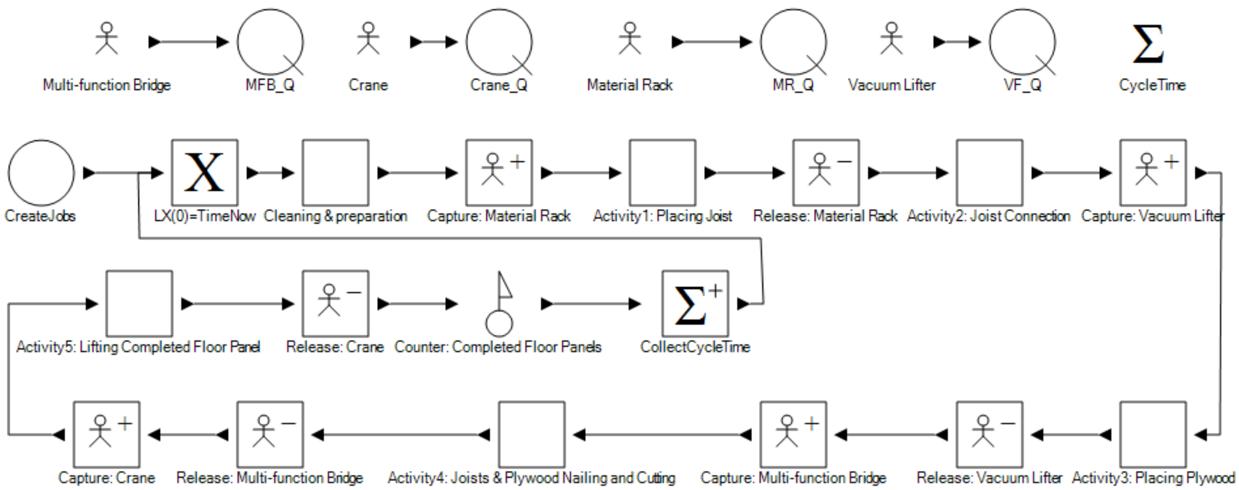

Figure 3: Simulation model of a floor workstation workflow

4. CASE STUDY

To verify the performance of the developed methodologies, this study takes an off-site construction factory in North America as an example. At this off-site construction factory, different building components, including walls, floors and roofs, are constructed in different workstations in parallel every weekday. Workers operate on a single shift from 8:00 AM to 5:00 PM each day, with a total of 8 working hours including a 1-hour lunch break. Six surveillance cameras are installed facing for two floor construction platforms and three roof construction platforms, from which real-time work conditions of machines and workstations can be captured.
Based on the manual observations and analysis of the surveillance videos from the case study, it's found that waiting for machines, particularly those shared across multiple workstations, was the most frequent and impactful source of delay. This study focuses on the floor workstations (equipped with two cameras), where two teams of workers work independently and two types of large machines, overhead crane and multi-function bridge, are shared by the workstations. The overhead crane is used for lifting the completed floor panels and the multi-function bridge is for automatic nailing and cutting of the floor panels. They can move flexibly between the two floor construction stations to meet the demands of production. The rough layout of the interested workstations and the samples of the shared machines at the factory are displayed in Figure 4.

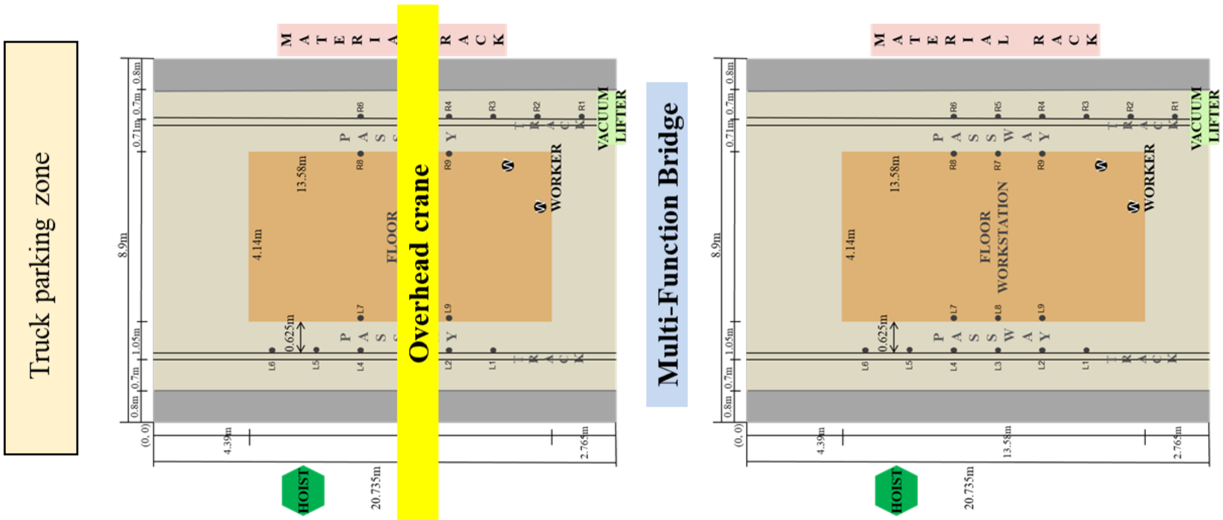

Figure 4: Interested workstations and machines in the off-site construction factory

The captured videos from the two cameras at floor workstations are inputted into the YOLO V8 model to automatically detect the interested machines and their coordinates in time series. Eight points at every interested construction platform are selected and their coordinates in real life are measured manually to compute the homography matrix with Equation [5]. The result of the homography matrix used in this case is shown in Equation 5, which is used to convert image coordinates into real-life coordinates. By integrating real-life coordinates with the detection-based tracking model introduced above, machine tracking and position mapping across multiple workstations are achieved. This study takes a whole-day construction process at the off-site construction factory as an example. Samples of the outputs generated from the computer vision model are displayed in Figure 5. Machine ID 1 represents the overhead crane and ID 2 refers to the multi-function bridge. The results are organized separately for each camera, while the object IDs keep consistent across cameras.

$$[5]\ H = \begin{bmatrix} 1.3981103 & 2.0567153 & -954.2132804 \\ -0.4156292 & 10.7977858 & -327.4255907 \\ -0.0001907 & 0.0040193 & 1.0000000 \end{bmatrix}$$

a) machine operation states:

Floor ID	Video Frame ID	Machine ID	X_center _real	Y_center _real	States
1	2049	2	308.0046	43.56003	Idle
1	2050	2	309.5867	43.51723	Idle
1	2051	2	307.5775	43.71439	Idle
1	2051	1	276.5522	63.12237	Transferring
1	2052	2	308.9794	43.42385	Idle
1	2052	1	268.8097	70.89481	Transferring
1	2053	2	308.9535	43.57717	Idle
1	2053	1	304.8615	77.23294	Transferring
1	2054	2	309.9313	44.1207	Idle
1	2054	1	289.969	81.93645	Transferring
1	2055	2	308.5341	43.48527	Idle
1	2055	1	306.8304	81.81741	Transferring
1	2055	1	319.0739	83.39898	Busy
1	2056	2	308.8568	43.17142	Idle
1	2057	2	310.5488	43.78712	Idle
1	2058	2	308.2364	42.54035	Idle
1	2058	1	341.3768	82.07899	Busy
1	2058	1	314.4478	82.8334	Busy
1	2059	2	308.461	43.69233	Idle
1	2059	1	343.9547	91.54151	Busy

b) machine requirements of a workstation:

Floor ID	Video Frame ID	Current time	Work condition	Machine Requirements
1	2048	08:34:08	connecting joist and panel and cutting extra panel	
1	2049	08:34:09	connecting joist and panel and cutting extra panel	Multi-function bridge
1	2050	08:34:10	connecting joist and panel and cutting extra panel	Multi-function bridge
1	2051	08:34:11	connecting joist and panel and cutting extra panel	Multi-function bridge
1	2052	08:34:12	connecting joist and panel and cutting extra panel	Multi-function bridge
1	2053	08:34:13	connecting joist and panel and cutting extra panel	Multi-function bridge
1	2054	08:34:14	lifting the completed panel	Overhead Crane
1	2055	08:34:15	lifting the completed panel	Overhead Crane
1	2056	08:34:16	lifting the completed panel	Overhead Crane
1	2057	08:34:17	lifting the completed panel	Overhead Crane
1	2058	08:34:18	lifting the completed panel	Overhead Crane
1	2059	08:34:19	lifting the completed panel	Overhead Crane
1	2060	08:34:20	lifting the completed panel	Overhead Crane
1	2061	08:34:21	lifting the completed panel	Overhead Crane
1	2062	08:34:22	lifting the completed panel	Overhead Crane
1	2063	08:34:23	lifting the completed panel	Overhead Crane
1	2064	08:34:24	lifting the completed panel	Overhead Crane
1	2065	08:34:25	lifting the completed panel	Overhead Crane
1	2066	08:34:26	lifting the completed panel	Overhead Crane
1	2067	08:34:27	lifting the completed panel	Overhead Crane

Figure 5: Samples of outputs from the computer vision-based model

Using the states and tracking information of each type of machine at multiple workstations, machine utilization rate is calculated with Equation [4]. Machine waiting time is computed by comparing the machine states (as listed in Figure 5a) with the machine requirements (as listed in Figure 5b) at the same frame ID using Equation [3]. Then the cumulative waiting time for each machine is obtained by adding the corresponding waiting time recorded by different cameras. In this case, the utilization rate of crane is around 36% and its waiting time occurring at multiple floor workstations is 8.23 minutes, taking less than 2% of daily working hours. The utilization rate of multi-function bridge is about 45% with a waiting time of 10.33 minutes in total for a whole day. The low values of utilization rate and machine waiting time demonstrate that the two types of shared machines among floor workstations are not fully used, showing high potential in developing new workstations in the off-site construction factory.

To improve machine utilization rate without adding extra long machine waiting time, different machine-workstation configurations are tested in the simulation environment. As shown in Figure 6, four workstations are built with the same SOP. Only the activities required shared machines are separately listed in the model, while other activities are roughly represented as preceding or succeeding activities. The production capacities of the existing workstations are learned from daily work schedules and computer vision analyzing results (see in Figure 4), which are inputted to the simulation model directly. Newly added workstations are assumed to have similar production capacity with the existing workstations so that similar a work schedule is followed.

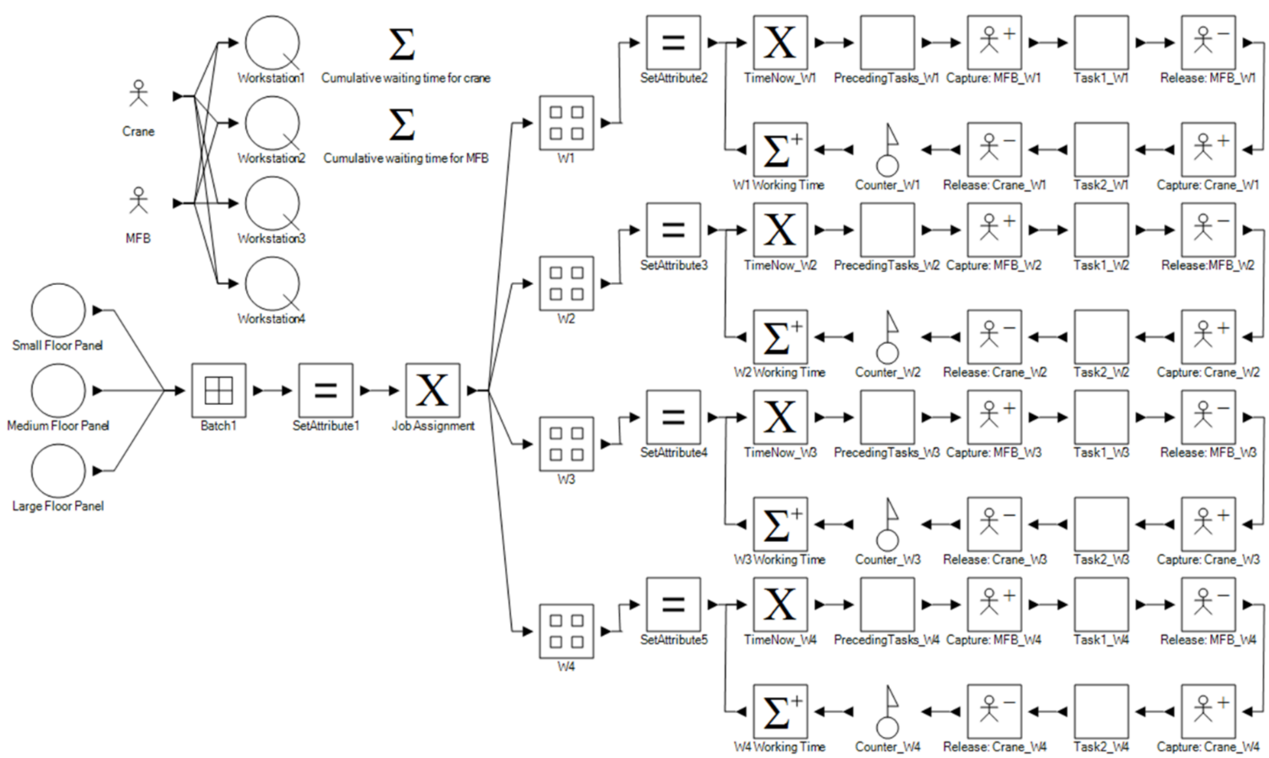

Figure 6: Simulation model with four workstations

The results for setting different numbers of workstations using the same shared machines are shown in Table 1, from which the best machine-workstation configuration, 4 workstations with one shared crane and one shared multi-function bridge, is found since more workstations will lead to an obvious increase in machine waiting time. Under the optimized machine-workstation configuration scenario, the utilization rate of overhead crane is around 72% and the multi-function bridge is about 81%, both improving more than 35% compared to the current configurations (2 workstations). These simulation results will not only be beneficial to the layout optimization of existing off-site construction factories but also be useful to guide the setup of a new off-site construction factory.

Table 1: Simulation-based machine-workstation configuration optimization

Number of workstations	Overhead crane		Multi-function bridge	
	Utilization rate	Waiting time	Utilization rate	Waiting time
2	36%	8.23 min	45%	10.33 min
3	53%	10.41 min	62%	13.67 min
4	72%	15.52 min	81%	19.66 min
5	89%	31.24 min	95%	37.93 min

5. CONCLUSIONS

This study proposes that machine utilization evaluation should account for both machine utilization rate and machine waiting time across multiple workstations to prevent construction delays caused by overloaded machine schedules. To automate the assessment of shared machine utilization, a computer vision framework is developed, integrating the YOLO V8 object detection model, the ByteTrack object tracking model, and a combination of position transition probability and appearance vectors for ID re-identification. This framework enables data collection across multiple machines and workstations. Machine operation states (idle, transferring, or busy) and workstation-specific machine demands are predicted based on computer vision results. These predictions are then used to evaluate machine utilization rates and waiting times for the current machine-workstation configuration. A low utilization rate suggests the potential for adding more workstations, provided that machine waiting times remain controlled. Conversely, excessive waiting times indicate the need for additional machines or a reduction in the number of workstations sharing a given machine. To determine the optimal configuration of shared machines and workstations, a simulation model replicating off-site construction processes is developed. This model evaluates the performance of various machine-workstation configurations, offering practical guidelines for setting up new off-site construction factories or optimizing work processes in existing ones. The proposed framework advances prior research by addressing critical limitations in both perception accuracy and dynamic optimization. By integrating hybrid tracking with real-time data-driven analysis, it achieves accurate machine-state recognition across multiple workstations. Compared to traditional work process monitoring systems, these innovations collectively demonstrate superior adaptability and cost-effectiveness, especially in scalable multi-workstation environments. Overall, this study makes three primary contributions: (1) evaluating and enhancing machine utilization from two critical perspectives—utilization rate and waiting time. (2) developing a machine work condition recognition framework using hybrid tracking-by-detection methods. (3) designing a simulation model to assess machine utilization under different machine-workstation configurations.

This research has several limitations that offer opportunities for future work. First, while the current study focuses on machine waiting time, waiting time also applies to workers, workstations, and material availability. Future research will expand the analysis to include these broader types of waiting time to enable a more comprehensive understanding of productivity constraints. Second, the current simulation model estimates task durations based on average values. To capture real-world variability, future work will integrate data-driven predictive models, such as regression analysis or time-series forecasting leveraging historical performance data, to estimate task durations more precisely. Lastly, the developed computer vision-based framework has been validated in a specific setting, but its generalizability can be further assessed by applying it to various production scenarios, including steel frame manufacturing, wooden roof construction, and onsite assembly involving mobile cranes. Also, to deal with the challenges in scaling the system to larger or more complex construction projects, other data collection tools, such as sensors and global position system, can be combinedly used in the developed object detection and tracking framework.

REFERENCES

AbouRizk, S., Hague, S., Ekyalimpa, R., and Newstead, S. 2016. Simphony: a next generation simulation modelling environment for the construction domain. *Journal of Simulation*, 10: 207–215.

Alsakka, F., El-Chami, I., Yu, H., and Al-Hussein, M. 2023. Computer vision-based process time data acquisition for offsite construction. *Automation in Construction*, 149: 104803.

Amiri, R., Sardroud, J. M., and Kermani, V. M. 2022. Decision support system for tower crane location and material supply point in construction sites using an integer linear programming model. *Engineering, Construction and Architectural Management*, 30: 1444–1462.

Arashpour, M., Wakefield, R., Blismas, N., and Lee, E. W. M. 2014. Analysis of Disruptions Caused by Construction Field Rework on Productivity in Residential Projects. *Journal of Construction Engineering and Management*, 140: 04013053.

Araya, F. 2022. Integration of discrete event simulation with other modeling techniques to simulate construction engineering and management: an overview. *Revista de la construcción*, 21: 338–353.

Chen, X., Wang, Y., Wang, J., Bouferguene, A., and Al-Hussein, M. 2024. Vision-based real-time process monitoring and problem feedback for productivity-oriented analysis in off-site construction. *Automation in Construction*, 162: 105389.

Dimaano, K. X. A., and Alon, A. S. 2023. Computer Vision-based Instance Segmentation and Data Augmentation for Construction Site Safety Compliance. *24th International Arab Conference on Information Technology (ACIT)*, Ajman, United Arab Emirates, 2023: 1-6.

Guo, S., Wang, S., Yang, Z., Wang, L., Zhang, H., Guo, P., Gao, Y., and Guo, J. 2022. A Review of Deep Learning-Based Visual Multi-Object Tracking Algorithms for Autonomous Driving. *Applied Sciences*, 12: 10741.

He, L., Zhou, Y., Liu, L., and Ma, J. 2024. Research and Application of YOLOv11-Based Object Segmentation in Intelligent Recognition at Construction Sites. *Buildings*, 14: 3777.

Innella, F., Arashpour, M., and Bai, Y. 2019. Lean Methodologies and Techniques for Modular Construction: Chronological and Critical Review. *Journal of Construction Engineering and Management*, 145: 04019076.

Jiang, Y., Li, M., Li, M., Liu, X., Zhong, R. Y., Pan, W., and Huang, G. Q. 2022. Digital twin-enabled real-time synchronization for planning, scheduling, and execution in precast on-site assembly. *Automation in Construction*, 141: 104397.

Karatas, I., and Budak, A. 2024. Deep learning-based recognition of construction activities in real construction site environment. *Engineering, Construction and Architectural Management*, 2024: 1036.

Li, J., Miao, Q., Zou, Z., Gao, H., Zhang, L., Li, Z., and Wang, N. 2024. A Review of Computer Vision-Based Monitoring Approaches for Construction Workers' Work-Related Behaviors. *IEEE Access*, 12: 7134–7155.

Li, J., Zhao, X., Kong, L., Zhang, L., and Zou, Z. 2024. Construction Activity Recognition Method Based on Object Detection, Attention Orientation Estimation, and Person Re-Identification. Buildings, 14: 1644

Li, G. Z., Xu, Z. G., Yan, S. L., Wang, H. Y., Bai, X. L., and Ren, Z. H. 2020. Bottleneck identification and alleviation in a blocked serial production line with discrete event simulation: A case study. *Advances in Production Engineering & Management*, 15.

Song, L., and Eldin, N. N. 2012. Adaptive real-time tracking and simulation of heavy construction operations for look-ahead scheduling. *Automation in Construction*, 27: 32–39.