INTEGRATING REAL-TIME OBJECT DETECTION INTO AN AR-DRIVEN TASK ASSISTANCE PROTOTYPE: AN APPROACH TOWARDS REDUCING SPECIFIC MOTIONS IN THERBLIGS THEORY

Xiang Yuan
Ph.D. Student, Department of Mechanical Engineering, University of Alberta, Canada

Qipei (Gavin) Mei
Assistant Professor, Department of Civil & Environmental Engineering, University of Alberta, Canada

Xinming Li
Assistant Professor, Department of Mechanical Engineering, University of Alberta, Canada

ABSTRACT: Due to challenges in filling vacant positions and the heightened demands posed on existing staff, employers and project managers are progressively considering the recruitment of inexperienced individuals and seeking strategies to swiftly provide them with essential job-specific knowledge. The potential of industrial AR has been widely researched to support workers in overcoming skill-related knowledge and enhancing industrial processes. However, most studies focus on demonstrating technology usability across different processes and overcoming engineering hurdles on a case-by-case basis. There is no direct benefit analysis on how AR assists construction tasks at human motion level, and how to eliminate the ineffective motions and reduce the duration of effective motions. To fill this gap, this paper first establishes an AR-based near real-time object detection system of small tools and components involved in task processes for egocentric perception of workers in the construction industry. Later, the Standard Operating Procedure (SOP) for scaffolding assembly activities is deconstructed from a manual process into Therbligs-based elemental motions. Finally, this research conducted a comparative study of two prototypes across four dimensions of evaluation. As a step forward in this direction, this paper renews the connotations of Therbligs theory under industry 5.0 era, rethinks the AR-assisted construction task processes, and applies appropriate technologies enhancing the adaptability of AR technology for construction workers' needs.

KEYWORDS: *Augmented Reality (AR); Microsoft HoloLens 2; Object Detection; Task Assistance; Therbligs*

1. INTRODUCTION

The construction industry, characterized as one of the least digitized and toughest labor challenges, urgently seeks solutions for industry transformation (Liao, Iseley, & Behbahani, 2022). In Canada, nearly half of the construction employers are facing the obstacle of recruiting skilled employees over the next three months (Government of Canada, 2022). Meanwhile, increased project complexity implies a high degree of technical, organizational and environmental variability and uncertainty, which leads to greater risk and poorer performance by construction personnel (Peñaloza, Saurin, & Formoso, 2020). In terms of technical requirements alone, project contributors must possess a more reliable and advanced technical qualification (Trinh & Feng, 2020). From the organizational and demographic characteristics, most construction workers rather rely heavily on previous work experience or oral interpretation from peers than follow the correct mechanical operation steps or standardized operation procedure (Ke, 2018). Oyekan et al., propose one of the mental challenges faced by workers, highlighting that as task operations become more complex, the cognitive load on operators also intensifies. This heightened cognitive load can be reflected in various human mental reactions of "Search," "Find," "Select," and other Therbligs. Consequently, due to the greater difficulty in filling vacant positions and the increased demands placed on employed personnel by the complexity of construction projects, employers and project managers are gradually hiring inexperienced newcomers and are eager to find countermeasures to equip them with the necessary job-related knowledge in a short time (Büttner, Prilla, & Röcker, 2020).

Industrial AR, deploying Augmented Reality (AR) technology into dynamic industrial environments targeting inhomogeneous user groups, is seen as potential solution to above dilemma and gains more momentum in both academic and industry (Grubert et al., 2010). The potential of Industrial AR has been widely researched to support workers in industrial scenarios in overcoming skill-related knowledge and enhancing industrial processes (de Souza Cardoso, Mariano, & Zorzal, 2020). In the context of industrial AR, one of the more fruitful and practical projects is the AR-Driven Task Assistance system, which supports workers by providing real-time sequence of assembly operations, tools to be used and collision free assembly paths at the workplace (Eswaran & Bahubalendruni, 2022). Previous research and the authors of this paper have also contributed to the development

Referee List (DOI: 10.36253/fup_referee_list)
FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup_best_practice)

Xiang Yuan, Qipei Mei, Xinming Li, *Integrating Real-Time Object Detection into an AR-Driven Task Assistance Prototype: An Approach Towards Reducing Specific Motions in Therbligs Theory,* pp. 121-132, © 2023 Author(s), CC BY NC 4.0, DOI 10.36253/979-12-215-0289-3.12

of a similar task-assisted AR prototype, which in this paper is expected to focus on the second function for highlighting recommended tools in the user's view while performing a task. Oyekan et al. addressed the mentioned challenge by using Therbligs to embed intelligence in workpieces and make them interactable and communicable. They developed smart workpieces to actively participate in assembly operations by providing their location and operational sequence to an operator (Oyekan et al., 2020). However, their solution may encounter various practical issues, such as overloaded servers when many workpieces update their status simultaneously, mispositioning of sensors and LEDs leading to electronic faults, and greater weight and more challenging manipulation of workpieces due to the addition of extra components.

It is obvious that CV and AR are theoretically linked and mutual fulfillment of each other, as OD could quickly identify and localize specific objects and draw bounding boxes around instances, and AR greatly extends users' capability and experiences by directly presenting detected objects and their digital data in an immersive, interactive way (Z. Wu, Zhao, & Nguyen, 2020). Thus, this paper proposes another computer vision-based solution to the same challenge by integrating OD into an AR-driven task assistance prototype. The choice of deployed technologies is highly related to their ability and referred to prior successful cases. While researchers have established the effectiveness of Industrial AR and its widespread adoption in construction task assistance over the past two decades, most studies focus on demonstrating technology usability across different processes and overcoming engineering hurdles on a case-by-case basis (Kim, Olsen, & Renfroe, 2022; S. Wu, Hou, Zhang, & Chen, 2022). However, user-related assessment of AR assistance systems and worker-oriented effectiveness in industrial environments is not a major focus (Tao, Lai, Leu, Yin, & Qin, 2019). To be more specific, there is no direct benefit analysis of how AR assists construction tasks at the human motion level and how to eliminate ineffective motions and reduce the duration of effective motions.

To fill this gap, this research first reports the further exploration of embedded object detection into the existing AR-driven task assistance prototype developed by authors. The existing AR prototype is targeted at construction workers without any previous work experience to conduct tasks from the beginning. But it only provides fixed information designed in advance about the activities and corresponding contents step by step. More advanced, the prototype developed in this research realizes a real-time detection of multiple scaffolding components, superimposes holographic texts, and gives hints about the correct selections which helps new industry entrants make the right choices from a wide range of tools and components. Later, the Standard Operating Procedure (SOP) of scaffolding assembly activity is decomposed from a human manual process into Therbligs-based elemental motions. It serves as both a specific example to enhance the understanding of Therbligs-based task processes and the foundation of subsequent benefit analysis. To present a more intuitive and clear effect, this research finally adopted a comparative study of a traditional AR prototype and an advanced AR prototype with object detection function from four dimensions of evaluation. It will demonstrate the superiority of the proposed prototype in easing cognitive load, eliciting contextual awareness, and reducing particular motion costs on Search, Select, and Find.

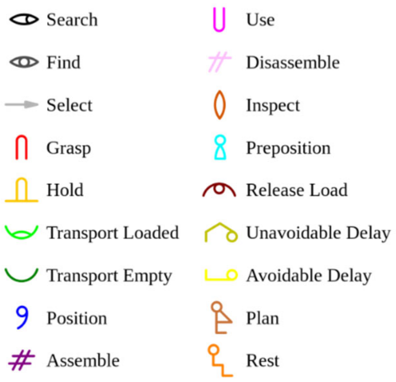
The proposed pathway not only explores the possibility of fully exploiting the advantages of both Augmented Reality and Object Detection, but also allows novice workers to easily perform high requirements tasks with a satisfied completion accuracy. As a step forward in this direction, this paper renews the connotations of Therbligs theory under industry 5.0 era, rethinks the AR-assisted construction task processes, and applies appropriate technologies enhancing the adaptability of AR technology for construction workers' needs. It is expected this research could inspire substantial discussions, enhance the implementation of AR-driven task assistance, and provide a valuable reference for construction workforce preparation.

2. RELATED WORK

2.1 Therbligs overview

Therbligs, first invented by Frank Gilbreth during the early 20th century, is a collection of 18 elemental human mental and physical motions used to describe any task and analyze the motion economy in the workplace (Sung, Ritchie, Lim, & Medellin, 2009). The full collection of 18 Therbligs and their symbols used for depicting when performing work is shown in Table 1. It is useful to use Therbligs to analyze the impact of technology adoption on individual earnings (Wang et al., 2021). The overall efficiency and productivity of tasks will be significantly improved because less time wasted on non-value-added activities and more time spent on productive work. The selection of Therbligs to be analyzed and addressed in this paper is not random. Taking consideration of the capabilities of computer vision technology and extended understanding of Therbligs connotation in the context of the construction tasks, this research mainly focuses on elemental motions of "Search", "Select" and "Find", which is also complied with previous research of Oyekan et al. The description of chosen Therbligs and their connotation in the context of the construction industry could be found in Table 2, gathering previous research and examples mentioned in the conversation with experts in both the ergonomics field and construction industry (David, 2000). For the application scenarios in this research, object detection function will be deployed to reduce the efforts needed by construction workers to search and select the tools and components they need.

Table 1: 18 Therbligs with symbols (Ninjatacoshell, 2012)



Search	Use
Find	Disassemble
Select	Inspect
Grasp	Preposition
Hold	Release Load
Transport Loaded	Unavoidable Delay
Transport Empty	Avoidable Delay
Position	Plan
Assemble	Rest

Table 2: Description of chosen Therbligs and their connotation in the context of the construction industry

Therbligs	Symbol	Description (Niebel & Freivalds, 2013)	Examples in the construction activities
Search	S	Eyes or hands groping for object; begins as the eyes move in to locate an object.	A construction worker looks for the location of a hammer in a warehouse.
Select	SE	Choosing one item from several; usually follows Search.	A construction worker selects the appropriately sized steel beam from a range of options.
Find	F	Defines the momentary mental reaction at the end of the Search cycle.	A construction worker realizes that he had found the correct 5 mm drill.

2.2 AR for Worker Onboarding and Skill Development

Industrial AR related to worker onboarding and skill development typically falls into two categories based on the research purposes and system functions: Step-by-Step Assistance AR and Hands-on Training AR (Butaslac, Fujimoto, Sawabe, Kanbara, & Kato, 2022). Both systems essentially start from the premise of breaking down knowledge barriers for people who do not have the ability or experience to perform the task contents. The difference between them is Hands-on Training AR will emphasize more on the knowledge stock after using the system and the ability to work independently when users are not equipped with system (Büttner et al., 2020), while Step-by-Step Assistance AR will emphasize the prompts, flexibility, and adaptive to users' needs and facilitate quicker familiarization and more regulated execution of predetermined task procedures (Zhang, Xuan, Yadav, Omrani, & Fjeld, 2023). For the user groups and specific scenarios targeted in this paper, the system built can be categorized as a Step-by-Step Assistance AR system.

2.3 Object Detection for AR

Numerous papers have extensively explored the applications of AR and OD within the construction industry, individually. From data preparation for construction objects, the traditional object detection dataset in the construction context is a collection of various categories (e.g., materials, workers, and their behavior of wearing PPE or falling from height), messy site layout, and large objects (e.g., heavy equipment of crane, excavators, bulldozers, and backhoe diggers). Thus, this research establishes a near real-time object detection dataset for small tools and components involved in task processes for workers' egocentric perception in construction industry. Besides, there exists relatively limited progress in cross-studies of AR and OD in this industry, despite its potential for significant advancement and promising opportunities. Wu et al. measured the utility and effectiveness of AR warning system on onsite construction workers with object detection for tracking onsite workers' locations and dynamic hazard areas (S. Wu, Hou, & Chen, n.d.).

Meanwhile, several other industries, such as smart manufacturing, have conducted successful research on the integration of AR and OD, which can serve as valuable points of reference for further exploration in the context of construction industry. They highlighted possible computing pathways, which could be broadly categorized based on where the data are handled into server-side processing, running locally on the device, or both (Ghasemi, Jeong, Choi, Park, & Lee, 2022). Considering the trade-offs between computational requirement and device limitation, cost and latency tolerance, and network connectivity, this research will adopt Microsoft Azure Custom Vision library as it offers a complete high-level solution suiting for HoloLens computing capabilities and is more common for implementation (Łysakowski et al., 2023).

Several publications investigated the utilization of Microsoft HoloLens 2 along with Azure Custom Vision services for object detection for different purposes (PatrickFarley, 2023). George created a training dataset of 215 images for motherboard and RAM in computer assembly task and reached an 80% match score even in varying environments, but also reveals challenges in limited experimental sample of three participants and false positives in similar components (George, 2021). Fuglseth created a proof of concept program for specific objects recognition and text information visualization in users' view with an open-source Microsoft COCO dataset (Fuglseth, 2022). Although this research demonstrated the technical feasibility of general objects detection in daily life, it lacks a specific use-case context and highlights the limitations of single object detection at a time. Casano used 9 specific classes of the COCO dataset and successfully implemented Azure Custom Vision object detector in the HoloLens for assisting and supporting users for better life or easier work style (Casano, 2021). This paper introduced a more mature and customized system by integrating eye motions, gestures, and voice commands, but faced limitations of predictions efficiency and more rigorous evaluation. Their pipelines to realize the object detection function are similar with each other. HoloLens will take a picture based on users' commands and uploads the picture to the Azure Custom Vision API. After successfully identifying, a label or images will be placed in users' AR view for easier awareness. These studies underscore the growing significance and feasibility of object detection in AR applications, point out challenges associated with their specific applications, and also highlight the potential for further improvements. It includes realizing a real-time object detection, testing its usability in practical application scenarios of industrial tasks, and developing a more powerful AR system to support more reliable multi-objects detection results.

3. PROTOTYPE DESIGN AND DEVELOPMENT

3.1 Prototype Overview

The fundamental idea behind the envisioned prototype involves the utilization of HoloLens 2 as an aiding instrument for promptly identifying objects in near-real time. Its primary function is to aid workers in identifying the precise tools and components required for the ongoing task phase. The target users for this prototype comprises generic individuals who lack prior work experience but seek rapid acquaintanceship with task-related details to ensure adherence to standards. For example, a novice construction worker aims to efficiently select tools and components aligned with the day's designated task and make high-quality commitments to their work activities. The object detection function will be triggered by users' voice commands, touch buttons, or gestures. The objects will be identified according to the current task step and its mentioned tools or parts. Once related objects are successfully recognized, the list of expected results will be three main parts (Farasin, Peciarolo, Grangetto, Gianaria, & Garza, 2020). A bounding box is a rectangle area that represents the object and its region. The class is a tag of the most probable category that the object belongs to. The probability score is the confidence level of algorithms in the detection accuracy and serves as a critical criterion for accepting or rejecting results. All information included will be displayed in the AR view as a visualized cue to workers.

3.2 Hardware and Software

Trimble XR10 with HoloLens 2 - Full Brim Hardhat is an integrated device in which a construction hard hat ensures easier wireless use in safety-constrained environments and the HoloLens 2 is the most commonly used XR headset. Microsoft HoloLens 2 is an ideal platform with high-tech hardware features for computer vision research, and also provides scalability of cloud services and connection to Microsoft Azure AI platform (Ungureanu et al., 2020). It sets a suitable equipment base that could serve multiple roles in proposed research and subsequent research, such as the source for capturing data in the form of video and frames, a computer of executing detection functions, and the tool for visualizing processed data and related task information (Qin et al., 2023). The proposed prototype is developed in the Unity cross-platform graphics engine (version 2022.3.3f1 LTS) using C# as programming language and MRTK (Mixed Reality ToolKit) packages for assets and interactive UI creation. Currently, this research adopted Microsoft Azure Cognitive Services to deploy object detection function by using

REST APIs and client library SDKs.

3.3 Object Detection using Azure Custom Vision in HoloLens

Microsoft Azure Custom Vision services enable users to rapidly customize cloud-based computer vision models and simply manage it using REST API calls. The overall design of architecture is shown in Figure 1. This research created the computer vision project in Azure Custom Vision portal and labelled a total of 12 classes including 1,224 images and 2,008 objects. The training model and its performance is further illustrated in both Section 4.4 and Section 5. The AR application, developed in Unity, adopted MRTK for user interaction and established a connection to the Azure Custom Vision API through endpoints. The Azure platform authenticates the AR system using Azure credentials and provides external GPU computational capabilities to preprocess the images and send the response through Wi-Fi. Once HoloLens receives object detection results from Azure Custom Vision, display holograms or annotations in the user world to indicate the detected objects.

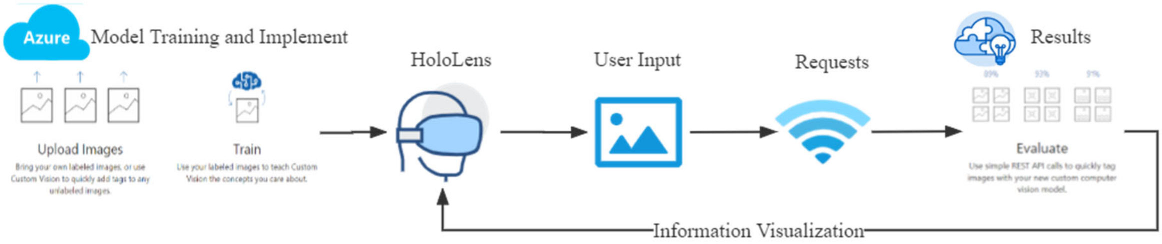

Azure
Model Training and Implement
Upload Images
Bring your own labeled images, or use Custom Vision to quickly add tags to any unlabeled images.
Train
Use your labeled images to teach Custom Vision the concepts you care about.
HoloLens
User Input
Requests
Results
Evaluate
Use simple REST API calls to quickly tag images with your new custom computer vision model.
Information Visualization

Figure 1 Pipeline design of using Azure Custom Vision and HoloLens 2

4. EXPERIMENT DESIGN FOR EVALUATION

The purpose of designing experiments is to provide a comprehensive insight into whether different forms of AR and varying technologies involved might impact user performance on the motion level. This study is designed with two independent variables: the complexity of construction tasks and the assisted tools used by participants to accomplish these tasks. Each variable is further subdivided, as task complexity has two levels (referred to as Task 1 and 2 hereafter), and using task-assisting tools comes in two types. Task 1 of the Miter Saw Stand assembly is the most complicated to build due to all the extra pieces and steps participants need to follow to make it work properly. Task 2 of the Scaffolding assembly is straightforward to understand what the task entails, but it's also easy to assemble it incorrectly and skip steps on some safety details. Detailed descriptions of these task specifications can be found in the subsequent section (section 4.3). The first type of assisted tools is a conventional AR prototype, which presents participants with guided text, images, and videos (referred to as Prototype 1 hereafter). On the other hand, the second type employs a more advanced AR prototype that includes object detection functionality to highlight crucial components for users (referred to as Prototype 2 hereafter).

4.1 Hypotheses

This research is formulated following hypotheses: **H1:** When using Prototype 2, participants were able to make fewer mistakes and complete the task with higher quality. **H2:** When using Prototype 2, participants were able to complete the task more efficiently, spending less time on "Search", "Select", and "Find" Therbligs. **H3:** When using Prototype 2, participants' cognitive demands were lower, and they can obtain better understandability for task contents and unfamiliar tools. **H4:** When using Prototype 2, participants think it is more intuitive, efficient, and enjoyable to use.

4.2 Bias Control

Potential bias and the effects of irrelevant factors, such as participants' familiarity with AR concepts and interaction, existing skills or learning curve, are more or less to interfere with the experiment results. The counterbalancing design principle and within-subject principle are throughout the entire experimental design and adopted controlling measures are stated as follows. Given that participants might have varying levels of proficiency with AR, some individuals are experienced users and developers, while others have a more superficial understanding [21]. Though researchers prepare an illustrative ppt for introducing experiments, explaining the meanings of prototype UI and panel, and briefly showing how to perform the sample tasks using different prototypes, researchers concern 2D-based explanation is less intuitive than real experience with device. Therefore, the prototype provides a comprehensive quick start guide to build perceptual awareness of AR capabilities and narrow knowledge gap among all participants. To mitigate the potential order effects and learning curve, there is also a clear definition of how to assign the participants into groups and decide their starting sequence. All participants will be randomly divided into two Group A and B. Both Group A and Group B will be exposed to two kinds of prototype and operate the same task, assembly Miter Saw Stand in the test phase 1 and assembly Scaffolding in the test phase 2, as shown in Figure 2. The difference between the two groups is Group A will start using Prototype 1 first, then they will shift to Prototype 2 in the test phase 2, while Group B will use two prototypes in reverse order.

4.3 Task Specification and Therbligs-based Information Presentation

As the basis of experiment design, this research selected the Metaltech Multipurpose 4-in-1 6 ft. Baker Scaffold as the specific user case for experiments, as the scaffolding is a typical and normal task in the construction site with higher hazards (Khan, Saleem, Lee, Park, & Park, 2021). Rigorously building up scaffolds is a vital safety management measure and prevents potential serious individual accidents. In terms of task content design, it can also be converted into another form of miter saw stand, which is a routine task for carpentry but different from those steps and used components in scaffolding assembly. Meanwhile, the difference in difficulty levels between the two tasks makes this choice more suitable for designing an experiment. After field assembly by three researchers, they agreed that the miter saw stand was the most complex and difficult to build, while the remaining three types were not as difficult to distinguish. As mentioned in previous section, the effect of decreased physical strength on experimental effectiveness, as well as increasing the difficulty differentiation between two tasks, two researchers worked together to simplify the task of scaffold assembly. Subjects would neither assemble the upper ladder to the lower part to prevent a total shelf height of about two meters, nor would they rotate the assembled shelf up and down. In the step of transitioning between the scaffold and the miter saw stand, they won't flip the entire platform due to its potentially harmful weight and surface area.

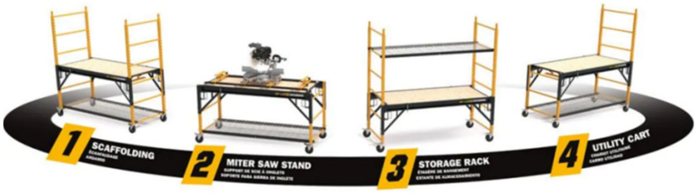


Figure 2 Selected Multipurpose 4-in-1 6 ft. Baker Scaffold for Experiment (The Home Depot, 2022)

Figure 3 shows the partial sequence of the assembly scaffold in the form of "Search", "Select", and "Find" Therbligs. The experiment workplace setup is shown in Figure 4, where all components are lying on the ground and a nearby shelf is for PPE equipment. "Search" Therbligs is reflected in locating the same type of tools from numerous building components in the package, such as searching ladders, braces, and locking pins. "Select" Therbligs happens less often than search, which is embodied in choosing the right one from a variety of similar things or alternatives, such as selecting a lower ladder from ladders where the upper ladder is a misleading option for participants (as shown in Figure *5*). "Find" Therbligs is a momentary mental activity that is reflected in the participant starting to move on to the next activity, such as the participant grabbing the searched brace and locking it in place using the U-lock kit.

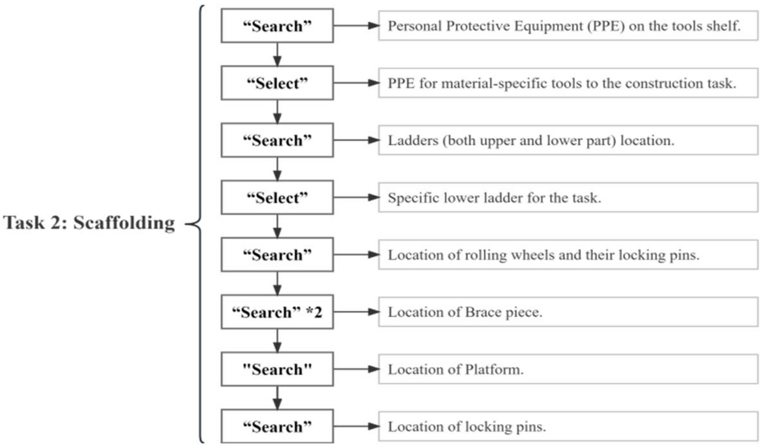

Task 2: Scaffolding
"Search"
Personal Protective Equipment (PPE) on the tools shelf.
"Select"
PPE for material-specific tools to the construction task.
"Search"
Ladders (both upper and lower part) location.
"Select"
Specific lower ladder for the task.
"Search"
Location of rolling wheels and their locking pins.
"Search" *2
Location of Brace piece.
"Search"
Location of Platform.
"Search"
Location of locking pins.

Figure 3 Partial sequence of assembly scaffold in the form of "Search", "Select", and "Find" Therbligs.

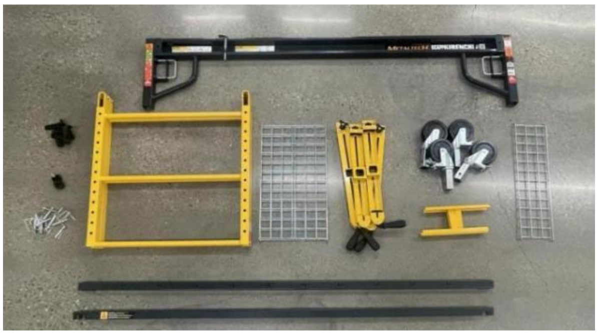


Figure 4 Experiment workplace setup.

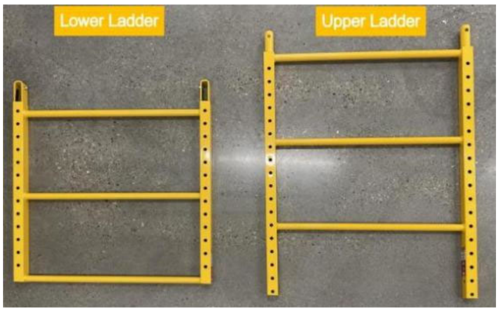


Figure 5 Experiment setting example for “select”.

4.4 Dataset Preparation

High-quality image data for target objects matters a lot to create a robust object-detection model (Lee, Jeon, & Shin, 2023). As shown in Table 3 and Table 4, this present study constructed a dataset to train and test the detection model, which took a total of 1,224 images, including 2,008 objects with 12 categories of classes. The dataset is compiled by two researchers using four devices of Hololens, iPhone, iPad, and Android Phone. This collection covers a diverse range of angles, lighting conditions, and backgrounds, drawing from the environments of two distinct research laboratories. In addition to sufficiently high-quality images of the objects in question, another important thing is the quality and quantity of annotation. The two researchers agreed on the labeling to ensure that the bounding box was strictly around each object. When one researcher’s annotation is complete, another researcher will cross-review each annotation result to ensure consistency. The quantity of each tag is roughly above 120 images, which makes the distribution even and not biased.

Table 3 List of scaffolding components for object detector training

No.	QTY. Used in EXP.	Class	No. Annotated Images
1	2	Lower ladder	137
2	1	Platform	144
3	4	Mounting bracket	194
4	2	Piece support	132
5	2	Brace	147
6	2	Shelf brace	169
7	1	Wire grid shelf (S)	216
8	5	Wire grid shelf (L)	217
9	4	5 in. caster	168
10	10	Locking pin	168
11	2	Anti-tip assembly	183
12	2	Tightening knob	128

Table 4 Examples of Captured Images

1. Lower ladder

2. Platform

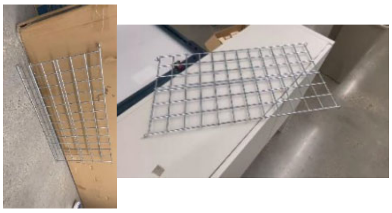
3. Mounting bracket

4. Piece support

5. Brace

6. Shelf brace

7. Wire grid shelf (S)


8. Wire grid shelf (L)

9. 5 in. caster

10. Locking pin

11. Anti-tip assembly

12. Tightening knob

4.5 Procedure and Evaluation Metrics

The overall experiment procedure is shown in Figure 6 an estimated duration of 20-25 minutes. Three GoPro cameras, each at an angle of 120 degrees to each other, were used to record the entire experiment to facilitate the subsequent analysis of the participants' Therbligs-based movements. Each participant will be through four steps: one preparation step, two test steps, and one after-testing step. After both Test Phase 1 and Test Phase 2, participants will be required to fill in a designed after-testing questionnaire immediately to express their direct subjective perception. At the end of the experiment, participants were allowed to volunteer for a short interview to express their opinions on improvements to the system prototype, their willingness to accept the technology, and other feelings not covered in the questionnaire. The data acquisition is based on a four dimensions evaluation: Quality, Efficiency, Mental Demand, and User Experience. The purpose of quality assessment is to detect the degree of precision in the work of the participants. This is done manually by experimental observers and is scored based on a complete error protocol. The error protocol describes errors in detail for two tasks, scores each of the two archetypes, and assigns three levels of scores according to the severity of the error (Wolf et al., 2021). The number of errors and the weighted total error score were finally statistically analyzed. Efficiency can be assessed in two ways. On a macro level, the overall time spent by each participant in completing the tasks using the two prototypes will be compared. On a micro level, experimental observers will use a timer to calculate the duration spent on the motion level of each Therbligs. This paper will explore to what degree user-centered process-oriented object detection has a significant effect on "Search", "Select", and "Find" Therbligs. Mental demand and user experience are mainly obtained through validated questionnaires and supplemented with optional semi-interviews.

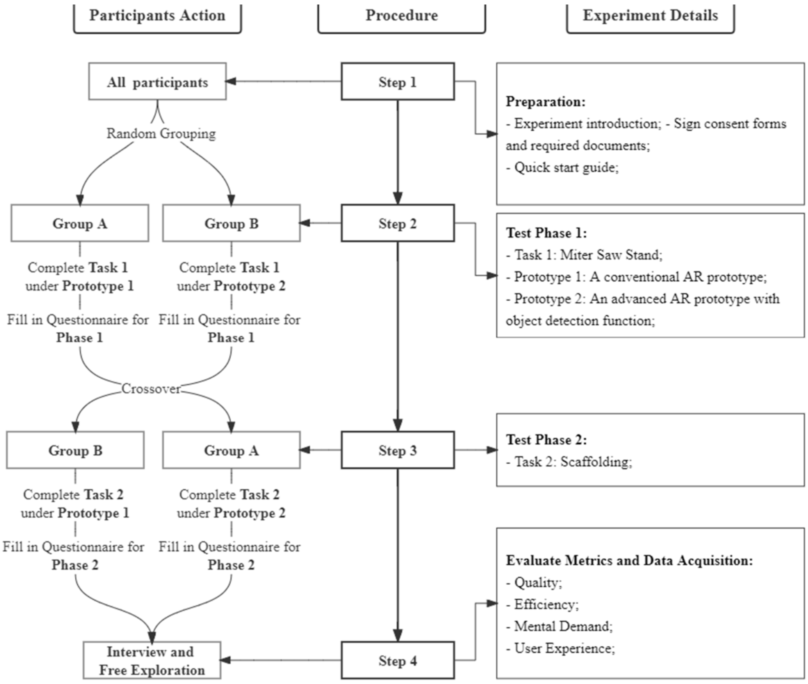
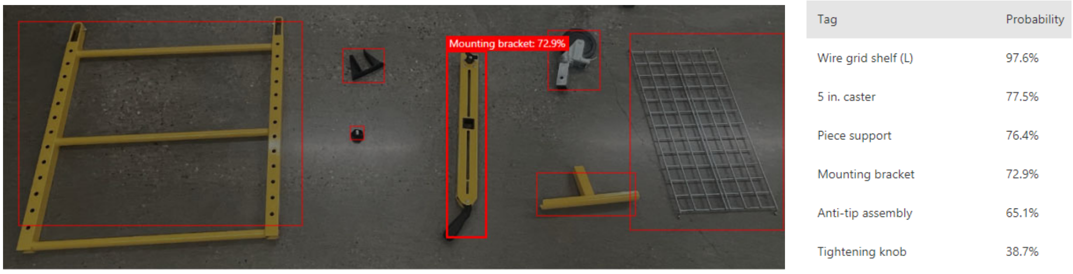

Participants Action
Procedure
Experiment Details
All participants
Random Grouping
Group A
Group B
Complete Task 1 under Prototype 1
Complete Task 1 under Prototype 2
Fill in Questionnaire for Phase 1
Fill in Questionnaire for Phase 1
Crossover
Group B
Group A
Complete Task 2 under Prototype 1
Complete Task 2 under Prototype 2
Fill in Questionnaire for Phase 2
Fill in Questionnaire for Phase 2
Interview and Free Exploration
Step 1
Step 2
Step 3
Step 4
Preparation:
- Experiment introduction; - Sign consent forms and required documents;
- Quick start guide;
Test Phase 1:
- Task 1: Miter Saw Stand;
- Prototype 1: A conventional AR prototype;
- Prototype 2: An advanced AR prototype with object detection function;
Test Phase 2:
- Task 2: Scaffolding;
Evaluate Metrics and Data Acquisition:
- Quality;
- Efficiency;
- Mental Demand;
- User Experience;

Figure 6 Experimental comparison between conventional AR and advanced AR with object detection

5. OBJECT DETECTION MODEL TRAINING RESULT

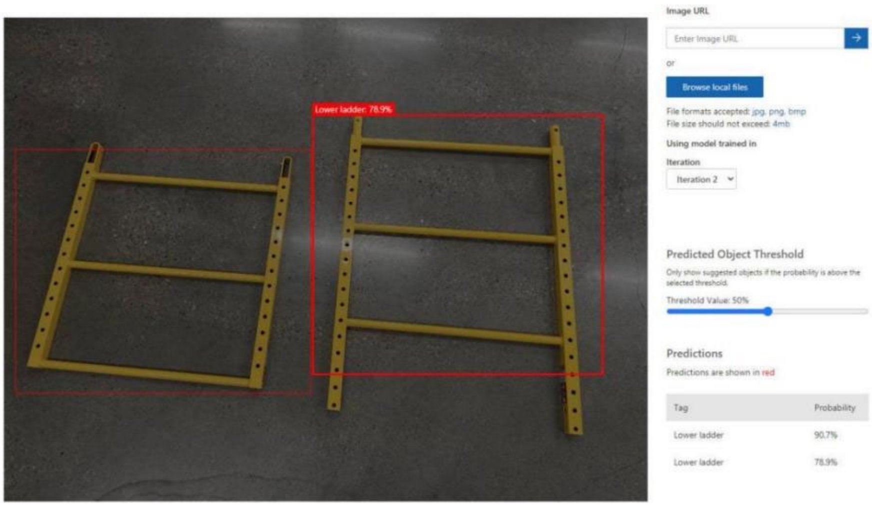
After training the model using 3 hours budget with General (compact) domain, the second-iteration training ended with 85.6% precision, 86.4 % recall, and 92% mAP. It is noted that when trying to train for more budget hours, the results remained the same which indicates that object detection model reaches its limitation by using current dataset. These metrics provide critical insights to evaluate the accuracy and effectiveness of object detection models. Precision is how many of the predicted instances are to be actually correct, recall gauges how well the model is capturing all the relevant correct instances, and mAP (mean Average Precision) represents the overall performance. This proposed model has a relatively higher performance in the mAP, which means that the model achieves a good balance between the precision and recall across different thresholds.

The excellent performance of this model is not only reflected in the data metrics, but also in the test images, as shown in . All objects, including very small Tightening Knob and Anti-tip Assembly objects, were successfully recognized one by one with a high success rate of more than 50%. However, there is still room for improvement in this model, as shown in Figure 8. As we mentioned above, we used the upper ladder as a misleading option, allowing participants to select the correct one from the two ladders for subsequent experiments. The trained model was unable to effectively distinguish between the two ladders when recognizing similar ones.



Tag	Probability
Wire grid shelf (L)	97.6%
5 in. caster	77.5%
Piece support	76.4%
Mounting bracket	72.9%
Anti-tip assembly	65.1%
Tightening knob	38.7%

Figure 7 Examples of Object Detection Results from developed model



File formats accepted: jpg, png, bmp
File size should not exceed: 4mb

Using model trained in

Iteration: Iteration 2

Predicted Object Threshold

Only show suggested objects if the probability is above the selected threshold.

Threshold Value: 50%

Predictions

Predictions are shown in red

Tag	Probability
Lower ladder	90.7%
Lower ladder	78.9%

Figure 8 Improvements needed in the model.

6. CONCLUSION AND FUTURE WORK

This research further developed the AR-Driven Task Assistance prototype by integrating object detection function to reduce elemental motions of “Search”, “Select” and “Find” in the Therbligs theory, and designed experiments to verify its direct benefits for construction workers and ease cognitive demanding during performing tasks. By integrating real time object detection into an AR-driven task assistance prototype, it is expected to enhance construction workers’ perception and situational awareness with a wearable, hands-free AR headset, which won’t interfere with workers’ current activities and enable a relatively larger and flexible Field of Vision (FoV) than mobile phone or tablets (Łysakowski et al., 2023).

This research is limited to a single pipeline to realize proposed application scenario, which leaves a researchable question on “Is there an optimal solution for the same function”. Since existing methods do not discriminate well between similar things, it is worth further improving the algorithm or exploring other publicly known algorithms in this domain. Besides, though there are some publications realizing a similar function on HoloLens, it is still worth comparing the performance of different algorithms by using a dataset of the same quality, diversity, and complexity. What’s more, this research is proposed to deploy real-time object detection into an AR-driven task assistance prototype and also verified by scaffolding and miter saw assembly activity, which is aimed at solving practical problems faced by construction industry. However, despite construction industry, other industries might encounter similar issues and challenges awaiting to be further improved. This leaves future efforts to generalize this proposed pathway to other industries and slightly adjust to their specific challenges.

REFERENCES

Butaslac, I. M., Fujimoto, Y., Sawabe, T., Kanbara, M., & Kato, H. (2022). Systematic Review of Augmented Reality Training Systems. *IEEE Transactions on Visualization and Computer Graphics*, 1–20. https://doi.org/10.1109/TVCG.2022.3201120

Büttner, S., Prilla, M., & Röcker, C. (2020). Augmented Reality Training for Industrial Assembly Work—Are Projection-based AR Assistive Systems an Appropriate Tool for Assembly Training? *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–12. Honolulu HI USA: ACM. https://doi.org/10.1145/3313831.3376720

Casano, D. (2021). *HoloHelp: HoloLens Detection for a Guided Interaction*. University of Catania.

David, D. (2000) Therbligs: The Keys to Simplifying Work, The Gilbreth Network: Therbligs. Available at: https://gilbrethnetwork.tripod.com/therbligs.html (Accessed: 26 July 2023).

de Souza Cardoso, L. F., Mariano, F. C. M. Q., & Zorzal, E. R. (2020). A survey of industrial augmented reality. *Computers & Industrial Engineering*, *139*, 106159. https://doi.org/10.1016/j.cie.2019.106159

Eswaran, M., & Bahubalendruni, M. V. A. R. (2022). Challenges and opportunities on AR/VR technologies for manufacturing systems in the context of industry 4.0: A state of the art review. *Journal of Manufacturing Systems*, *65*, 260–278. (36). https://doi.org/10.1016/j.jmsy.2022.09.016

Farasin, A., Peciarolo, F., Grangetto, M., Gianaria, E., & Garza, P. (2020). Real-time Object Detection and Tracking in Mixed Reality using Microsoft HoloLens: *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, 165–172. Valletta, Malta: SCITEPRESS - Science and Technology Publications. https://doi.org/10.5220/0008877901650172

Fuglseth, S. S. (2022). *Object Detection with HoloLens 2 using Mixed Reality and Unity a proof-of-concept* (Bachelor thesis, Høgskolen i Molde - Vitenskapelig høgskole i logistikk). Høgskolen i Molde - Vitenskapelig høgskole i logistikk. Retrieved from https://himolde.brage.unit.no/himolde-xmlui/handle/11250/3023916

George, R. (2021). *Using Object Recognition on Hololens 2 for Assembly* (M.S.). Retrieved from https://www.proquest.com/docview/2621280160/abstract/34D5D9DBA2ED4869PQ/1

Ghasemi, Y., Jeong, H., Choi, S. H., Park, K.-B., & Lee, J. Y. (2022). Deep learning-based object detection in augmented reality: A systematic review. *Computers in Industry*, *139*, 103661. https://doi.org/10.1016/j.compind.2022.103661

Government of Canada, S. C. (2022). Retrieved from https://www.statcan.gc.ca/en/subjects-start/labour_/labour-shortage-trends-canada#shr-pg0

Grubert, J., Hamacher, D., Mecke, R., Böckelmann, I., Schega, L., Huckauf, A., … Tümler, J. (2010). Extended investigations of user-related issues in mobile industrial AR. *2010 IEEE International Symposium on Mixed and Augmented Reality*, 229–230. https://doi.org/10.1109/ISMAR.2010.5643581

Ke, Y. (2018). Research on the Chinese Industrialized Construction Migrant Workers from the Perspective of Complex Adaptive System: Combining the Application of SWARM Computer Simulation Technology. *Wireless Personal Communications*, *102*(4), 2469–2481. https://doi.org/10.1007/s11277-018-5266-8

Khan, N., Saleem, M. R., Lee, D., Park, M.-W., & Park, C. (2021). Utilizing safety rule correlation for mobile scaffolds monitoring leveraging deep convolution neural networks. *Computers in Industry*, *129*. Scopus. https://doi.org/10.1016/j.compind.2021.103448

Kim, J., Olsen, D., & Renfroe, J. (2022). Construction Workforce Training Assisted with Augmented Reality. *2022 8th International Conference of the Immersive Learning Research Network (iLRN)*, 1–6. https://doi.org/10.23919/iLRN55037.2022.9815960

Lee, K., Jeon, C., & Shin, D. H. (2023). Small Tool Image Database and Object Detection Approach for Indoor Construction Site Safety. *KSCE Journal of Civil Engineering*, *27*(3), 930–939. https://doi.org/10.1007/s12205-023-1011-2

Liao, W., Iseley, T., & Behbahani, S. (2022). *Industry/University Cooperative Research Centers (IUCRC): A Critical Component for Addressing Underground Infrastructure Challenges*. 56–66. https://doi.org/10.1061/9780784484289.007

Łysakowski, M., Żywanowski, K., Banaszczyk, A., Nowicki, M. R., Skrzypczyński, P., & Tadeja, S. K. (2023, June 6). *Real-Time Onboard Object Detection for Augmented Reality: Enhancing Head-Mounted Display with YOLOv8*. arXiv. Retrieved from http://arxiv.org/abs/2306.03537

Niebel, B., & Freivalds, A. (2013). *Niebel's Methods, Standards, & Work Design*. McGraw-Hill Education.

Ninjatacoshell. (2012). *English: The 18 therbligs.* Own work. Retrieved from https://commons.wikimedia.org/wiki/File:Therblig_(English).svg

Oyekan, J., Hutabarat, W., Turner, C., Arnoult, C., & Tiwari, A. (2020). Using Therbligs to embed intelligence in workpieces for digital assistive assembly. *Journal of Ambient Intelligence and Humanized Computing*, *11*(6), 2489–2503. https://doi.org/10.1007/s12652-019-01294-2

PatrickFarley. (2023, July 18). What is Custom Vision? - Azure AI services. Retrieved October 10, 2023, from https://learn.microsoft.com/en-us/azure/ai-services/custom-vision-service/overview

Peñaloza, G. A., Saurin, T. A., & Formoso, C. T. (2020). Monitoring complexity and resilience in construction projects: The contribution of safety performance measurement systems. *Applied Ergonomics*, *82*, 102978. https://doi.org/10.1016/j.apergo.2019.102978

Qin, Y., Wang, S., Zhang, Q., Cheng, Y., Huang, J., & He, W. (2023). Assembly training system on HoloLens using embedded algorithm. *Third International Symposium on Computer Engineering and Intelligent Communications (ISCEIC 2022)*, *12462*, 121–128. SPIE. https://doi.org/10.1117/12.2660940

Sung, R. C. W., Ritchie, J. M., Lim, T., & Medellin, H. (2009). *Assembly planning and motion study using virtual reality*. 31–38. Scopus. https://doi.org/10.1115/WINVR2009-713

Tao, W., Lai, Z.-H., Leu, M. C., Yin, Z., & Qin, R. (2019). A self-aware and active-guiding training & assistant system for worker-centered intelligent manufacturing. *Manufacturing Letters*, *21*, 45–49. https://doi.org/10.1016/j.mfglet.2019.08.003

The Home Depot. (2022). Retrieved from https://www.homedepot.ca/product/metaltech-scaffold-bench-multipurpose-4-in-1-6-ft-baker-scaffold/1001160246

Trinh, M. T., & Feng, Y. (2020). Impact of Project Complexity on Construction Safety Performance: Moderating Role of Resilient Safety Culture. *Journal of Construction Engineering and Management*, *146*(2), 04019103. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001758

Ungureanu, D., Bogo, F., Galliani, S., Sama, P., Duan, X., Meekhof, C., … Pollefeys, M. (2020, August 25). *HoloLens 2 Research Mode as a Tool for Computer Vision Research*. arXiv. https://doi.org/10.48550/arXiv.2008.11239

Wang, B., Zhang, Z., Jiang, C., Zhao, Y., Ding, S., Xu, F., & Niu, J. (2021). A Novel Approach Combined with Therbligs and VACP Model to Evaluate the Workload During Simulated Maintenance Task. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12771 LNCS*, 164–173. Scopus. https://doi.org/10.1007/978-3-030-77074-7_13

Wolf, J., Wolfer, V., Halbe, M., Maisano, F., Lohmeyer, Q., & Meboldt, M. (2021). Comparing the effectiveness of augmented reality-based and conventional instructions during single ECMO cannulation training. *International Journal of Computer Assisted Radiology and Surgery*, *16*(7), 1171–1180. https://doi.org/10.1007/s11548-021-02408-y

Wu, S., Hou, L., & Chen, H. (n.d.). *Measuring the impact of Augmented Reality warning systems on onsite construction workers using object detection and eye-tracking*.

Wu, S., Hou, L., Zhang, G. (Kevin), & Chen, H. (2022). Real-time mixed reality-based visual warning for construction workforce safety. *Automation in Construction*, *139*, 104252. https://doi.org/10.1016/j.autcon.2022.104252

Wu, Z., Zhao, T., & Nguyen, C. (2020). 3D Reconstruction and Object Detection for HoloLens. *2020 Digital Image Computing: Techniques and Applications (DICTA)*, 1–2. https://doi.org/10.1109/DICTA51227.2020.9363378

Zhang, Y., Xuan, Y., Yadav, R., Omrani, A., & Fjeld, M. (2023, February 3). *Playing with Data: An Augmented Reality Approach to Interact with Visualizations of Industrial Process Tomography*. arXiv. https://doi.org/10.48550/arXiv.2302.01686