Vision-based 6-DOF Grasp Method in Cooperative Robotic Wood Fabrication

Taking a reciprocal Christmas Tree as an example

Yalei Zhu[1], Tianyi Gao[2], Quanpi Zheng[3],Philip F. Yuan[4]*
[1,2,3,4]Tongji University
[1,2,3,4] {archzylei|2211046|2411915|philip007}@tongji.edu.cn

The circular reuse of irregular natural timber offers ecological promise for architecture but faces technical barriers in automated construction. To address the geometric variability of reclaimed logs in the context of cooperative robotic wood fabrication, this study proposes a vision-based 6-DOF grasping method. It is implemented in construction within a cooperative robotic wood fabrication framework. The method synergizes visual identity with multi-objective optimization, enabling the fabrication of a 2.8-meter reciprocal structure through three integrated phases: (1) Vision-based log recognition and localization, (2) Prior grasp pose optimization with milling feasibility, and (3)grasp pose matching within the spatial scene. This method achieves vision-based 6-DOF grasping of non-standard geometries, replacing traditional manual measurement and wood fixing procedures. It effectively adapts to natural curvature and surface irregularities of the logs, significantly improving fabrication efficiency. This research establishes a scalable pathway for repurposing non-standard timber at architectural scales, advancing sustainable construction through the fusion of circular material practices and robotic fabrication.

Keywords: *Robotic Fabrication, Natural Log, 6-DOF Grasp, Multi-objective Optimization*

INTRODUCTION

Amid rising environmental concerns and the push for carbon reduction, low-carbon renewable materials—especially natural wood—are gaining attention in sustainable construction. While reused wood offers strong potential due to its renewability and versatility (Sormunen and Kärki, 2019), its irregular geometry makes automated processing difficult. Unlike standardized timber, which benefits from mature robotic workflows for cutting and assembly, natural logs require labor-intensive manual operations in critical fabrication stages—grasping, positioning, and machining—due to their non-uniform shapes and variable surface features. This reliance on human intervention not only increases production costs but also limits the scalability of natural timber applications in modern architecture.

To address this challenge, we propose a vision-based 6-DOF grasping workflow to replace manual operations. The system consists of: (1) depth camera–based recognition and localization of irregular logs, (2) multi-objective optimization to predict 6-DOF grasp poses in advance, and (3) digital twin–driven scene matching and execution of grasping using a dual-robot setup. This approach replaces manual trial-and-error with algorithmic prediction, reducing pre-machining log fixation time by 77% while ensuring rational grasp configurations and achieving sub-centimeter accuracy in downstream fabrication. This method is applied in the construction of a Reciprocal Christmas Tree made from 28 irregular logs. The project demonstrates the vision-based system's ability to recognize, localize, and accurately predict suitable grasp poses. This research advances the automation of irregular material processing and offers a scalable solution for integrating natural logs into sustainable construction.

LITERATURE REVIEW

Digital fabrication of natural wood

The digital design and fabrication of natural timber generally involve several key stages, including material collection, information acquisition, allocation design, and experimental construction. However, the process remains challenged by the inherent irregularity of the material. Common methods for material data acquisition include 3D scanning (Self, 2016), while processing stages utilize MR (Lok and Bae, 2022), CNC (Larsson et al., 2019), and robotic arms (Geno et al., 2022). (Saslawsky et al., 2021) outlined a five-step workflow from scanning to assembly, and (Chai, 2024) explored collaborative robotic fabrication for forked logs. Despite these advances, many steps still rely on manual work, limiting scalability. To address this, recent research emphasizes automation. Autonomous construction robots with on-site sensors (Kim et al., 2019) enable real-time feedback. For example, (Melenbrink et al., 2020) combined LiDAR scanning with 2D imaging for geometric control, and (Aleksandra Anna Apolinarska et al., 2021) used six-axis force/torque sensors to monitor timber assembly in real time. However, current literature offers limited discussion on automated workflows specifically tailored to natural logs.

Automatic recognition and localization

Image-based deep learning models, combined with depth cameras, enhance object detection accuracy and robustness. Recent studies demonstrate that 3D data from depth cameras optimizes detection performance. For instance, Avola et al. (2022) proposed a dual-branch neural network with depth cameras for industrial defect detection. Yuan et al. (2024) improved localization using optimized YOLOv5s and depth sensors. Ahmad and Rahimi (2022) highlighted that depth data mitigates limitations of 2D image processing in complex environments, enhancing model stability. While effective in cluttered scenes, these methods focus on isolated detection and lack integration with downstream fabrication. They also struggle to fuse multi-modal data like machining positions for unified decision-making.

6-DOF Task-Oriented Grasping

Replacing manual gripper control with automated recognition and grasping is pivotal. Manual grasping involves measuring and marking irregular surfaces to identify nodes and select grasp points (Chai, 2024), followed by collision simulation—a time-consuming and labor-intensive process. In this context, 6-DOF category-level task-oriented grasping is proposed. This technique enables robotic end-effectors to control both position (X, Y, Z) and orientation (roll, pitch, yaw), dynamically generating grasp strategies tailored to object categories (e.g., timber) and tasks (e.g., milling). Current autonomous grasping methods (Huang et al., 2022; Hundt et al., 2019; Zeng et al., 2022) analyze object topology via deep learning to infer collision-free grasps, yet are limited to simple tasks (e.g., pick-and-place). However, in robotic wood fabrication, milling forces generate strong torque on logs, making traditional two-finger grippers insufficient. A rail-mounted dual-gripper system is needed to resist torque and ensure stability during machining. This requires redefining grasp pose prediction as a gripper-log matching task.

METHOD

This study aims to: (1) develop an automated method combining deep learning–based vision, 6-DOF autonomous grasping, and robotic wood fabrication; (2) optimize robotic performance for complex geometries; and (3) explore the architectural use of irregular timber through advanced robotics. The proposed vision-based 6-DOF grasping method includes three steps: (1) recognizing and localizing irregular logs using YOLOv8 and depth images, (2) predicting candidate grasp poses with MOEA/D, and (3) selecting the optimal pose using digital twin technology (Figure 1).

Each log is scanned and reconstructed, with the YOLOv8 model applied to depth camera images for autonomous recognition and spatial localization. A set of 6-DOF candidate grasp poses is predicted using the MOEA/D. These poses are registered with the reconstructed model in the scene, and after evaluation, the optimal grasp pose is selected for execution, enabling subsequent cooperative robotic wood fabrication.

Dataset setup

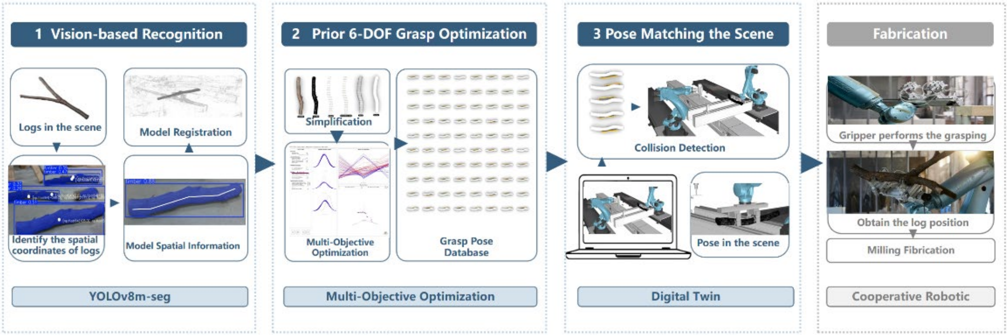
The quality of the dataset for deep learning-based object detection and localization models determines the number of recognizable categories and the accuracy of results. In the experiments, natural logs must be identified and localized within the processing environment. Data samples are classified based on image features. Training and validation are conducted using distinct batches of data samples, and the trained weight files are deployed on the same batch of logs to evaluate recognition outcomes.

Under uniform lighting conditions simulating the workspace, we captured 300 images of camphor wood logs (with bark), containing approximately 900 logs photographed from various angles. These images were annotated on the CVAT platform to generate segmentation masks for each log. The images and their corresponding annotation files collectively form the dataset for model training.

Recognition Algorithm

This study focuses on bridging the recognition gap in automated processing workflows through computer vision, rather than improving existing algorithms. Thus, the mature YOLOv8 framework is adopted. YOLOv8 (You Only Look Once Version 8), developed by Ultralytics, (Jocher, Chaurasia, & Qiu, 2023) is a real-time object detection model. Its architecture integrates the CSPDarknet53 backbone, Anchor-Free detection head, and dynamic decoupled head, minimizing computational redundancy . It is widely applied in industrial inspection, autonomous systems, and natural resource management.

Figure 1
Workflow Diagram



Implementation for Natural Log Segmentation

Using YOLOv8m-seg, we trained an instance segmentation model for camphor wood logs. The model detects bark-covered logs in scene images, extracts masks, and performs instance segmentation. The system integrates an Intel RealSense D435i RGB-D depth camera, which features 1280 × 720 @ 90 fps depth output, a stereo IR setup with an 87° × 58° field of view, and depth accuracy better than 2% at 2 m, to capture depth at bounding-box centers or key points. These values are then converted via camera intrinsics into 3D spatial coordinates for localization.

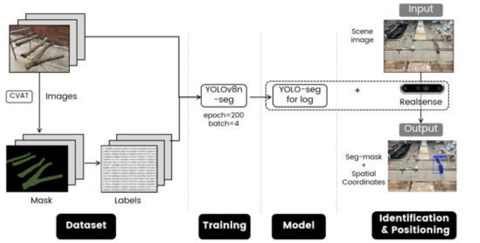


Figure 2
Performance of log instance segmentation model based on YOLOv8-Seg

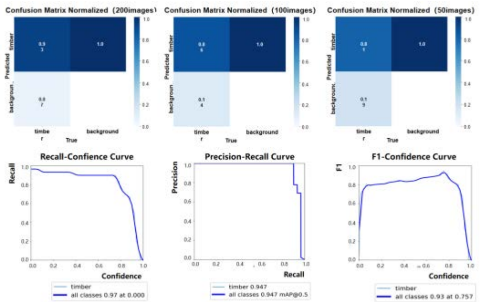


Figure 3
Effectiveness of YOLOv8-Seg-ased Log instance segmentation model

Training Configuration

The model was trained for 200 epochs with an input image size of 640×400 pixels and a batch size of 4 images. Each training session lasted approximately 10 minutes, achieving efficient optimization without compromising segmentation accuracy. Implemented using Python and PyTorch, the training was conducted in the PyCharm environment. A hand–eye calibrated Realsense D435i was integrated to capture scene images. Post-training, the YOLOv8m-seg-timber model leverages pixel-level depth data from the camera to compute spatial coordinates, enabling precise 3D localization of logs in cluttered environments. This configuration ensures robust adaptation to real-world timber processing scenarios while maintaining computational practicality (Figure 2).

Recognition Result

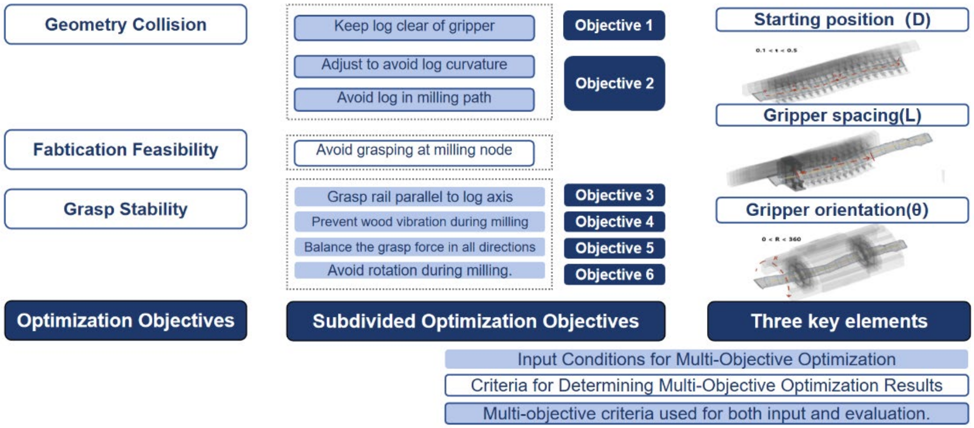
We selected the model trained on a dataset of 200 images for recognition and localization. The Loss value showed a convergent trend with increasing training epochs, proving that the model's recognition accuracy improved over time. The final model achieved an accuracy of 97%, recall of 94%, and F1 score of 93% (Figure 3).

The MOEA/D Model for Predicting 6-DOF Grasp Poses of Natural Logs

The prediction of 6-DOF grasp poses is the second step in the vision-based grasping method. By frontloading the computationally intensive multi-objective optimization, the overall workflow becomes more streamlined and efficient.

Given the multiple factors affecting grasping in cooperative robotic wood fabrication, MOEA/D (Multi-Objective Evolutionary Algorithm based on Decomposition) is adopted as the optimization strategy. It considers three core objectives: geometric collision avoidance, grasp stability, and processing feasibility. Within this framework, these objectives are further broken down into nine sub-objectives—six used as input parameters for optimization, and three as screening criteria to evaluate results. The inputs are quantified to optimize three grasp parameters: starting point, jaw distance, and jaw angle (Figure 4). Final grasp poses are selected through collision checks, ensuring both stability and processing suitability.

Figure 4
Optimization diagram



Objective 1: Maximizing the Average Distance Between Branches and Rails， i.e., minimize f_1 To avoid collisions between the gripper and the log, the goal is to minimize grasping protruding sections from bent logs. This is quantified by the distance between the branches and the rails. A larger protrusion reduces the distance, making the grasp less favorable. The optimization seeks to maximize the average distance between *n* points (15 cm apart) from the jaws to the rails, influenced by *d*, *L*, and *θ*.
The larger the average distance between the branches and the rails, the smaller $f_1(d, L, \theta)$.

$$f_1(d, L, \theta) = \frac{n}{\sum_{i=1}^{n} A_i B_i} \qquad (1)$$

Objective 2: Maximizing Node Processability, i.e., minimize f_2 To avoid collisions with the milling tool, the gripper should not approach the log from the direction of nodes. The closer the gripper's orientation is to a node, the less favorable the grasp. Each grasp pose is assigned a processability score F_i. for each node a_i, based on how much the gripper's non-processable area overlaps with the node. If there is direct overlap, the pose is considered invalid. For a log with n nodes, all scores F_i are combined into a weighted processability map to evaluate overall suitability.

Each grasp pose receives a node processability score from the distribution map, reflecting its suitability for milling. This objective is influenced by the parameters *d*, *L*, and *θ*.

$$f_2(d, L, \theta) = \frac{1}{\sum_{i=1}^{n} \omega_i F_i} \qquad (2)$$

The higher the node processability score, the smaller $f_2(d, L, \theta)$. However, this objective alone cannot guarantee avoidance of collisions between the milling tool and the log or rails.

Objective 3: Minimize the angle between the grasping claw normal vector and the log centerline tangent vector, i.e., minimize f_3 To ensure both claws grasp the log stably, the grasping rail should align with the log's centerline. A larger angle between them reduces grasp stability, while a smaller angle improves it. This is measured by the angle α between the claw's normal vector $\vec{A}$ and the log's tangent vector $\vec{B}$, using the absolute value of cosine similarity to account for directional uncertainty. A higher cosine value indicates better alignment and a more favorable grasp.

$$f_3(d,L,\theta) = \frac{1}{|cos\,\alpha|} = \left|\frac{|\vec{A}||\vec{B}|}{\vec{A}\cdot\vec{B}}\right| = \left|\frac{\sqrt{\sum_{i=1}^{n}\vec{A_i}^2}\ \sqrt{\sum_{i=1}^{n}\vec{B_i}^2}}{\sum_{i=1}^{n}\vec{A_i}\times\vec{B_i}}\right| \quad (3)$$

Objective 4: Minimize Vibration and Displacement Caused by Cutting Forces, i.e., minimize f_4 Milling-induced chatter can increase processing errors. To ensure machining accuracy, the grasping pose should be adjusted to minimize vibration and displacement caused by cutting forces. This can be formulated as follows:

$$f_4(d,L,\theta) = \sqrt{{\delta_x}^2 + {\delta_y}^2 + {\delta_z}^2} = \sqrt{\left(\frac{F_x\cdot d}{E\cdot A}\right)^2 + \left(\frac{F_y\cdot d}{E\cdot A}\right)^2 + \left(\frac{F_z\cdot d}{G\cdot J}\right)^2} \quad (4)$$

The smaller the vibration and displacement caused by cutting forces, the smaller $f_4(d,L,\theta)$

Objective 5: Maximize Gripping Force Uniformity, i.e., minimize f_5 The gripping stability is measured by the uniformity of the gripper force distribution in three directions. The objective function for the force uniformity is given by:

$$f_5(d,L,\theta) = -\frac{min(R_{g1}x,R_{g2}x)}{max(R_{g1}x,R_{g2}x)} - \frac{min(R_{g1}y,R_{g2}y)}{max(R_{g1}y,R_{g2}y)} - \frac{min(R_{g1}z,R_{g2}z)}{max(R_{g1}z,R_{g2}z)} \quad (5)$$

The higher the gripping force uniformity, the smaller $f_5(d,L,\theta)$.

Objective 6: Minimize System Torque, i.e., **minimize f_6** To ensure machining accuracy, the gripper pose should be adjusted to minimize the wood's rotation caused by these forces, i.e., minimize the system's torque. The total system torque, based on equations (6), (8), and (9), is expressed as:

$$f_6(d,L,\theta) = M_{G_c} = M_P + M_{g1} + M_{g2} \quad (6)$$

Multi-Objective Optimization: To achieve an optimal solution, the normalized values of the six objective functions f_1, f_2, f_3, f_4, f_5, f_6 are combined in a multi-objective optimization model. The objective is to minimize the following function:

$$f(d,L,\theta) = \sum_{i=1}^{n}\omega_i\cdot f_i(d,L,\theta) \quad (7)$$

Where the total system torque is the sum of the milling force torque, gravity torque, and gripper torque. The smaller the total system torque, the smaller $f_6(d,L,\theta)$.

Where ω_1 to ω_6 represent the importance of each objective function and are implemented in the Grasshopper platform via MOEA/D (Multi-Objective Evolutionary Algorithm based on Decomposition). MOEA/D decomposes the multi-objective problem into multiple scalar subproblems. Through neighborhood-based information sharing, it efficiently searches the Pareto front and identifies diverse optimal grasp solutions (Chettibi, 2024)

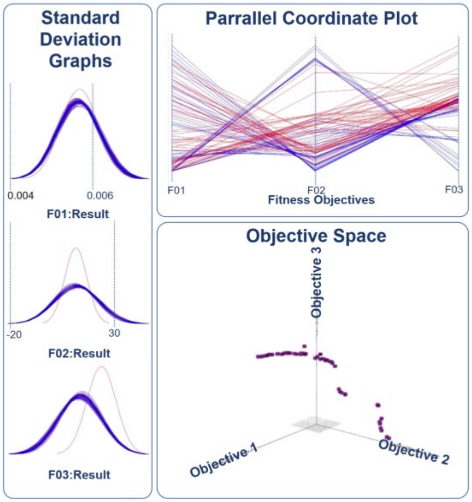


Figure 5
Multi-Objective Optimization Process

CONSTRUCTION PRACTICE

This study validates the proposed methodology through the construction of the "Reciprocal Christmas Tree," a 2.8-meter-tall conical structure composed of 28 irregular wooden logs sourced from storm-fallen trees in a forest park (Figure 6). It includes abandoned Metasequoia, pine, willow, and plane trees with an average diameter of 10 cm. The design workflow began by generating a structural axis system to define spatial relationships, followed by 3D scanning of every log to capture geometric variations and natural defects. The scanned data were matched with the theoretical axes to generate customized half-lap joints, ensuring precise adaptation to the irregular shapes of the reclaimed logs while maintaining structural continuity. At the same time, each component not only has a node in the middle, but the end position of the component also needs to be shaped and processed according to the Boolean calculation of the geometric body in the space (Figure 7).

In the phase of robotic fabrication, a dual-KUKA KR2700-210 system was coordinated via the proprietary FURobot platform (Lu et al., 2020) (Figure 8). The primary robot, equipped with a high-speed spindle (18,000 rpm), performed milling operations to carve 82 non-standard half-lap joints with linear profiles (maximum depth: 50 mm) (Figure 9). Concurrently, the secondary robot executed material handling tasks, utilizing an integrated vision system to identify and grasp logs based on real-time point cloud analysis. The FURobot platform enabled synchronized operation, dynamically adjusting machining paths and clamping orientations using UDP (User Datagram Protocol) communication. By incorporating adaptive feed rates (400–500 mm/min) aligned with wood grain density, the system achieved joint-fitting accuracy within ±0.5 mm. All 28 logs were processed within two days. During fabrication, the stable grasping success rate reached 71%. Under stable grasping conditions, milling tasks were completed collision-free, with task-relevant grasp success rate of 65% within tolerance thresholds.

The average fabrication time per log totaled 40 minutes (including automated recognition, grasping, positioning, and milling). Notably, the

Figure 6
Reciprocal Christmas Tree

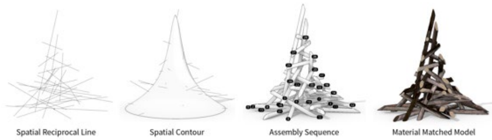


Figure 7
Design development

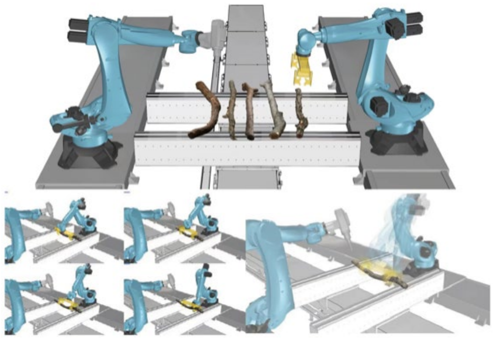


Figure 8
Dual robot platform

preparatory steps—"on-site registration, optimal grasp pose selection, and grasping implementation"—required an average of only 8 minutes per log. In contrast, the traditional manual workflow involving "manual measurement and marking of node positions, grasp pose prediction based on node locations, and manual feeding operations" averaged 35 minutes per log. This comparison demonstrates significant improvements in both efficiency and precision through automation.

Figure 9
Robotic wood joint fabrication

On-site assembly employed mixed reality (MR) guidance to streamline installation. Before on-site assembly, we use Fologram projected holographic overlays onto the real world, visually guiding workers through sequential assembly steps and highlighting joint engagement logic in advance (Figure 10). The MR system dynamically visualized spatial deviations and provided corrective feedback, enabling rapid positioning of the irregular logs. This approach facilitated the complete installation of the structure within 4.5 hours (Figure 11), with all half-lap joints exhibiting a maximum assembly gap of 2.4 mm. The holistic integration of multi-species reclaimed timber, dual-robot collaboration, and MR-assisted assembly demonstrates the robustness of the proposed framework for sustainable, irregular wood construction.

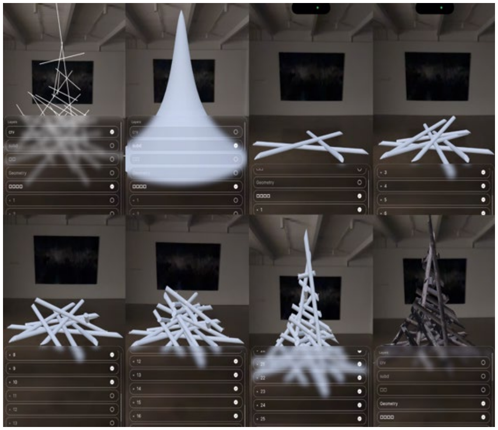


Figure 10
Mixed reality assembly simulation

CONCLUSION

This study presents an integrated robotic workflow that synergies deep learning with multi-objective optimization to enable autonomous grasping and machining of irregular natural logs. Validated through the construction of a 28-log reciprocal structure, the proposed system demonstrates significant advancements in automation efficiency: vision algorithms achieve robust log recognition despite bark occlusion, grasp optimization reduces manual preprocessing by 77%, and dual-robot collaboration ensures sub-centimeter milling

Figure 11
On-site assembly

precision for half-lap joints. By overcoming key bottlenecks in material variability and geometric unpredictability, the workflow successfully transforms irregular wood—often dismissed as construction waste—into architecturally viable components, aligning circular material reuse with industrial automation.

However, scalability challenges persist, particularly when processing logs with complex bifurcations or dense nodal clusters. Current limitations stem from unmodeled grasp interactions at acute angles and insufficient training data for highly variable bark textures. Future work will prioritize three advancements: First, developing bifurcation-aware grasping models to address force distribution imbalances and collision risks at angled nodes. Second, enhancing vision robustness through expanded datasets capturing diverse bark morphologies and environmental conditions. Third, implementing adaptive machining sequences that dynamically adjust toolpaths during multi-stage fabrication of clustered joints. These improvements aim to extend the framework's applicability to broader natural timber typologies, including branched and knotted specimens. By bridging the gap between ecological material practices and robotic adaptability, this research lays critical groundwork for scaling sustainable construction in the era of Industry 4.0.

REFERENCES

Avola, D., Cascio, M., Cinque, L., Fagioli, A., Foresti, G. L., Marini, M. R., and Ros-si, F. (2022). 'Real-time deep learning method for automated detection and localization of structural defects in manufactured products' *Computers & Industrial Engineering*, 172, 108512. https://doi.org/10.1016/j.cie.2022.108512

Chai, H. (2024). 'Integrated workflow for cooperative robotic fabrication of natural tree fork structures' *Automation in Construction*. https://doi.org/10.1016/j.autcon.2024.105524

Clark, C., Jambeck, J., and Townsend, T. (2006). 'A Review of Construction and Demolition Debris Regulations in the United States' *Critical Reviews in Environmental Science and Technology*. https://doi.org/10.1080/10643380500531197

De Wolf, C., Hoxha, E., and Fivet, C. (2020). 'Comparison of environmental assessment methods when reusing building components: A case study' *Sustainable Cities and Society*, 61, 102322. https://doi.org/10.1016/j.scs.2020.102322

D, A., C, B., & J, M. (1996). *Handbook of sustainable building. An environmental preference method for selection of materials for use in construction and refurbishment*. https://www.aivc.org/resource/handbook-sustainable-building-environmental-preference-method-selection-materials-use

Dean, B., Dulac, J., Petrichenko, K., and Graham, P. (2016). *Towards zero-emission efficient and resilient buildings: Global Status Report*. *Towards zero-emission efficient and resilient buildings*. Global Alliance for Buildings and Construction (GABC).

Geno, J., Goosse, J., Van Nimwegen, S., and Latteur, P. (2022). 'Parametric design and robotic fabrication of whole timber reciprocal structures' *Automation in Construction*, 138, 104198. https://doi.org/10.1016/j.autcon.2022.104198

Huang, X., Halwani, M., Muthusamy, R., Ayyad, A., Swart, D., Seneviratne, L., Gan, D., and Zweiri, Y. (2022). 'Real-time grasping strategies using event camera' *Journal of Intelligent Manufacturing*, 33(2), 593–615. https://doi.org/10.1007/s10845-021-01887-9

Hundt, A., Jain, V., Lin, C.-H., Paxton, C., and Hager, G. D. (2019). 'The CoSTAR Block Stacking Dataset: Learning with Workspace Constraints' *2019 IEEE/RSJ Inter-national Conference on Intelligent Robots and Systems*

(IROS), 1797–1804. https://doi.org/10.1109/IROS40897.2019.8967784

Larsen, N. M., Aagaard, A. K., Kjærvik, P. L., and Harðarson, M. K. (2024). Constructing with Discarded Wood. In: Eversmann, P., Gengnagel, C., Lienhard, J., Ramsgaard Thomsen, M. and Wurm, J. (eds), *Scalable Disruptors. DMS 2024*. Cham: Springer. https://doi.org/10.1007/978-3-031-68275-9_8

Larsson, M., Yoshida, H., and Igarashi, T. (2019). 'Human-in-the-loop fabrication of 3D surfaces with natural tree branches' *Proceedings of the ACM Symposium on Computational Fabrication*, 1–12. https://doi.org/10.1145/3328939.3329000

Lok, L., Samaniego, A., and Spencer, L. (2021). 'Timber de-standardized' *Proceedings of the 41st Annual Conference of the Association of Computer Aided Design in Architecture (ACADIA)*, New York, NY, USA, 25–30. https://www.academia.edu/download/117553125/acadia21_222.pdf

Melenbrink, N., Werfel, J., and Menges, A. (2020). 'On-site autonomous construction robots: Towards unsupervised building' *Automation in Construction*, 119, 103312. https://doi.org/10.1016/j.autcon.2020.103312

Sormunen, P. and Kärki, T. (2019). 'Recycled construction and demolition waste as a possible source of materials for composite manufacturing' *Journal of Building Engineering*, 24, 100742. https://doi.org/10.1016/j.jobe.2019.100742

Self, M. (2016). Hooke Park: Applications for timber in its natural form. *Advancing Wood Architecture*. Routledge.

Sui Pheng, L. (2001). 'Construction of dwellings and structures in ancient China' *Structural Survey*, 19(5), 262–274. https://doi.org/10.1108/02630800110412859

Teischinger, A., Krug, D., Sandberg, D., and Tobisch, S. (2023). Sawn-Timber Products. In: Niemz, P., Teischinger, A. and Sandberg, D. (eds), *Springer Handbook of Wood Science and Technology*. Springer Handbooks. Cham: Springer. https://doi.org/10.1007/978-3-030-81315-4_25

Torghabehi, O. O., Vliet, K., Von Buelow, P., and Mankouche, S. (2018). 'Limb: Inventory-constrained design method for application of natural tree crotches as heavy timber joinery' *TxA Emerging Design+ Technology Conference.* Fort Worth, Texas, USA.

Wen, B., Lian, W., Bekris, K., and Schaal, S. (2022). 'CaTGrasp: Learning Category-Level Task-Relevant Grasping in Clutter from Simulation' *2022 International Conference on Robotics and Automation (ICRA)*, 6401–6408.