Article | Received 1 September 2026; Revised 9 November 2026; Accepted 11 November 2026; Published 16 November 2026
https://doi.org/10.55092/sc20260021

Laser-triggered and vision-validated collaborative method for node recognition of rebar tying robots

Jingwen Liang[1], Yingdi He[1], Cheng Zhou[1], Yan Zhou[1,2,*], Shifeng Wen[2] and Xi Zeng[3]

[1] National Center of Technology Innovation for Digital Construction, School of Civil and Hydraulic Engineering, Huazhong University of Science and Technology, Wuhan, China
[2] State Key Laboratory of Material Processing and Die & Mould Technology, School of Materials Science and Engineering, Huazhong University of Science and Technology, Wuhan, China
[3] Department of Civil and Environmental Engineering, College of Design and Engineering, National University of Singapore, Singapore, Singapore

* Correspondence author; E-mail: yanzhou@hust.edu.cn.

Highlights:

- A task-level laser-triggered and vision-validated method is proposed.
- The method achieves a 92.0% mAP@0.5 and an 8.2% false-binding rate.
- The YOLO inference workload is reduced by 50.5% via laser-triggered frames.

Abstract: The accuracy of node perception in rebar tying robots determines the safety of the tying. The current laser-triggering-only solution cannot judge the state of the node, thus resulting in incorrect tying. Based on the multi-channel binary laser signal; this paper puts forward a collaborative recognition method that can be used for image collection and YOLOv5 classification: When the laser array is activated, a local image is obtained, and You Only Look Once version 5(YOLOv5) is then used to determine the category of the node; only nodes that are tieable will be tied. Based on the 60-image held-out test set, the model has an overall mean Average Precision at IoU=0.5 (mAP@0.5) of 92.0%; and the tieable node (normal_node) is 96.8%. A comparison of 91 physical nodes has shown that the false-binding rate of the above-mentioned method is only 8.2%; it is better than the laser-only scheme (45.1%) and close to that of the baseline using vision only. In addition, offline inference tests on the Intel Core i7 processor showed that the laser-triggered mode requires that only about 49.5% of the frames be processed; and thus the work required for YOLO inference has been reduced to about 50.5%. Without changing the existing hardware, the proposed method has increased the reliability of the decision to tie and provided a very cost-effective perception solution for engineering deployment.

Keywords: reinforcement bar tying robot; joint recognition; multimodal collaboration; laser triggering; YOLOv5

Copyright©2026 by the authors. Published by ELSP. This work is licensed under Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium provided the original work is properly cited.

1. Introduction

1.1. Research background and engineering requirements

In recent years, with the rapid development of smart construction technologies, construction robots have gradually become an important tool for improving construction efficiency and alleviating labor shortages [1,2]. Reinforcement tying is a crucial process in the construction of prefabricated buildings and cast-in-place concrete structures, and it has always been done by hand, resulting in high physical labour, harsh working environments and uneven quality. As building structures have grown more complex, the number and accuracy demands for tie joints have also risen [3].

Taking a typical cast-in-place concrete structure as an example [4], rebar tying generally includes three parts: workers first mark and place the rebar on the already laid base layer, then place the transverse and longitudinal rebars one by one in position, and finally use a tying hook to wrap and tighten the wire at every intersection. A good worker can tie about 100 to 150 joints per hour, but efficiency decreases significantly after 2–3 hours of continuous work. Figure 1 is a diagram of the manual tying process and some typical intersections of the rebar mesh.

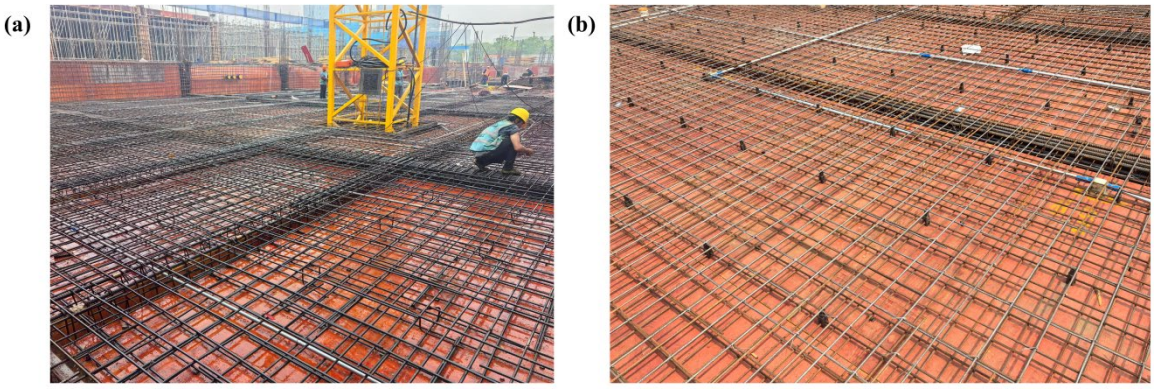


Figure 1. Schematic diagram of reinforcement tying operations and joints. **(a)** Manual tying operation; **(b)** Intersection of reinforcement mesh.

In terms of tying quality, it is difficult to ensure that manual tying is consistent - when workers are tired, quality problems such as missing ties, loose ties or insufficiently tight ties will occur [5]. They are more serious in a tight construction schedule or a large-scale operation, and thus affect the quality of project acceptance and structural safety.

More seriously, there will be a shortage of workers in the construction industry. According to the data from the National Bureau of Statistics, the average age of construction workers across the country will be over 48 years old in 2025 [6], and because there are no young people to replace them, the cost of labour has been increasing continuously. Given the above situation, the development of robotic systems that can automatically identify and tie rebar joints has begun to be an area of research in the field of smart construction [7]. Although robot tying has the advantage of being standardised, accurate joint detection is a basic requirement for engineering implementation [8]. Robots should be able to correctly perceive the spatial position and current state of the rebar intersection nodes to plan the path for tying, align the head of the tie, and carry out the tying action [9,10].

1.2. Major challenges in identifying reinforcement tie nodes

Identifying rebar tying joints presents the following three engineering challenges:

(1) The construction environment is complex, and visual perception is prone to interference. Factors such as drastic changes in lighting, metal reflections, and dust obstruction at construction sites lead to unstable image quality, causing the recognition performance of single-vision solutions to fluctuate significantly under complex conditions.

(2) Commonly used laser sensors provide limited output information. Single-point laser sensors, widely used in actual engineering applications, output only binary signals; they can only determine whether an object is present within the detection range but cannot distinguish the specific type of tie or its binding status.

(3) Tie states are diverse and require fine-grained discrimination. Ties can be in various states, including untied, tied, and abnormal; existing solutions struggle to provide sufficient information to support robots in making accurate tying decisions.

1.3. Scope and innovative aspects of this study

To address the complementary shortcomings of purely vision-based approaches—namely, their lack of stability—and purely laser-based approaches—namely, their inability to distinguish node states—this paper proposes a node recognition method that combines laser triggering with vision verification, designed for use in cost-sensitive construction scenarios. The core approach does not involve improving a specific detection algorithm, but rather designing a division of labor between vision and laser at the system architecture level: using laser signals as a "conditional activation switch" for the vision module to establish a "trigger-verification" task-level collaboration mechanism.

Specifically, when the horizontal and vertical laser arrays are triggered at the same time, the system determines that a possible node exists under the current tying head and immediately activates the You Only Look Once version 5 (YOLOv5) vision recognition module to perform inference on the local image beneath the tying head. The system will only send a "Tie" command when the model determines that the node is tieable; if it determines that the node has already been tied, is abnormal, or there is some interference, it will send a "Skip" command. The main contributions of this paper are as follows:

(1) We propose a task-level multimodal collaborative architecture based on "laser triggering + visual verification. Through the cooperation of the laser and the vision system, this improves the reliability of tying decisions without changing the existing hardware.

(2) By using laser signals as the conditional activation signal for the vision module, visual inference is only required to process the triggered frames (with a trigger rate of approximately 49.5% in the test scenario), thereby engineering the "event-driven" perception strategy for rebar tying applications.

(3) We have carried out tests in the actual environment of rebar mesh and have shown that the collaborative mechanism works well; we also quantified the actual contribution of visual inspection to reduce the rate of false binding and improve accuracy.

It should be noted that the proposed method is mainly designed for horizontal surface-layer rebar meshes. At this time, the tying head is at a certain height from the plane of the rebar mesh and can work

properly. Due to the sensor trigger logic and fixed camera position for multiple-layer rebar meshes with different heights or other types of 3D rebar cages with different directions, it is restricted in application.

2. Related work

2.1. Reinforcement binding robot and joint detection system

Research on rebar-tying robots has mainly focused on designing the main body of the robot, motion planning, and developing end-effectors. Based on the architecture of the perception system, there are three types of existing methods [11,12]; they are shown in Figure 2.

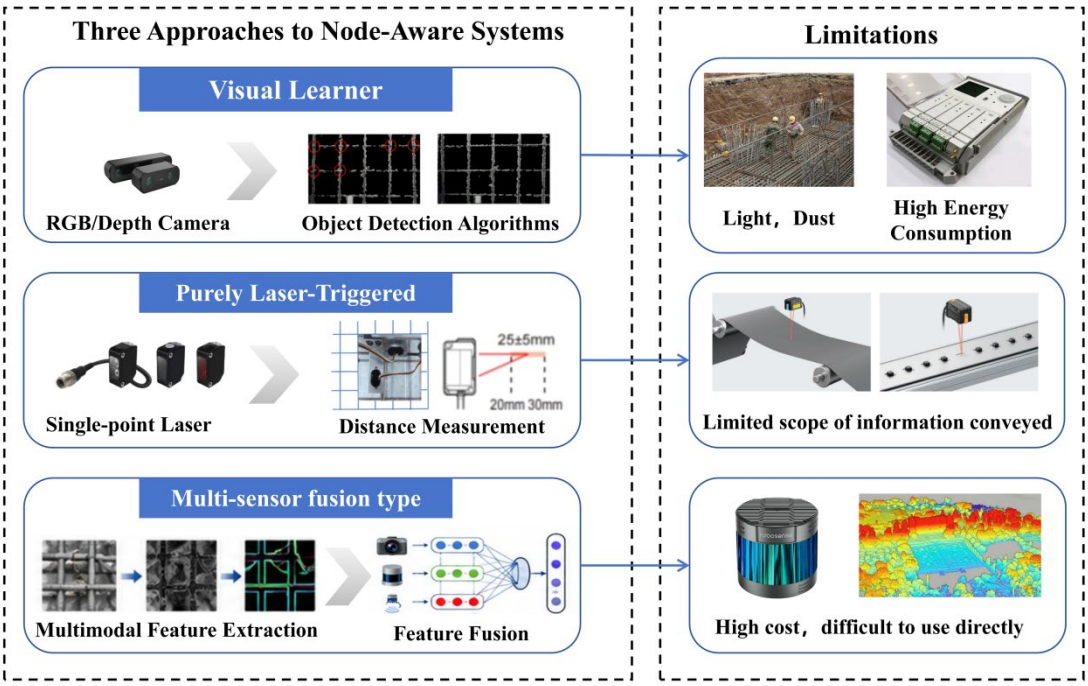


Figure 2. Schematic diagram of classification of reinforcement joint sensing solutions.

The first type is the architecture of vision-based design. The system uses an Red, Green, Blue (RGB) camera as the only sensory input and detects the areas where rebar intersects using object detection algorithms [13], and guides the tying head to perform the necessary actions. Some research has tried to add depth cameras to get 3D information [14], but the perception logic is still "continuous acquisition—continuous inference," and the vision sensors are always working, resulting in a large amount of computation.

The second type is a pure laser-triggered system. A single-point laser sensor is placed near the tying head, and it emits a binary signal to determine if there is an object in the detection range [15]. This architecture features fast response times and is unaffected by lighting conditions; however, since the sensor output provides only "present/absent" information, it cannot distinguish the specific type of node or the tying status, resulting in a high rate of incorrect tying [16,17].

The third type is the multi-sensor fusion architecture. The above systems combine visual and laser range information at the feature or decision layer; they are currently used in autonomous driving and industrial robotics [18]. However, in the case of rebar tying, because of cost and installation space

restrictions, laser sensors are mostly single-point rather than scanning types, so it is difficult to obtain continuous point cloud data, and existing fusion methods cannot be directly applied.

2.2. Visual and multimodal detection methods for reinforcement joints

The main task of rebar joint inspection is to find out the spatial position of the rebar intersections, so as to give a reference point for tying. Based on image analysis, visual localisation can be used to find the intersection points of rebars. Early research was based on traditional image processing methods; that is, using the Hough transform to detect straight lines in the parameter space of edge points by voting in the Hough space [19]; The LSD algorithm directly extracts the endpoints of the line segment [20]; and the EDLines algorithm improves the real-time performance of detection based on this method [21]. These methods are highly sensitive to image quality; edge detection will be unstable under different lighting conditions, reflections, or occlusions [22,23], and they can only obtain two-dimensional intersection coordinates and cannot provide node classification information or tie status.

With the advancement of deep learning technology, object detection algorithms based on convolutional neural networks have been widely applied to rebar joint recognition. The YOLO series of algorithms, with their single-stage detection mechanism and real-time performance advantages, are extensively used for rebar joint detection [24–26]; two-stage detectors such as Faster Region-based Convolutional Neural Network (Faster R-CNN) also demonstrate excellent detection performance [27]. Li *et al.* [28] classified tie point states into two categories—tied and untied—and employed an improved YOLOX-s network to detect tie point states, combining it with depth maps to locate 3D coordinates, thereby achieving high detection accuracy in precast concrete (PC) component scenarios. However, this method belongs to a vision-dominated architecture—where visual sensors operate continuously—resulting in low computational resource utilization in scenarios with sparse joints. Given the cost sensitivity of construction projects, hybrid solutions based on low-cost hardware are more economically feasible in practical engineering applications.

To solve the problems of the current research, researchers have tried to use Light Detection and Ranging (LiDAR) data to develop multimodal detection methods [29]. Existing fusion strategies can be divided into two types: feature-level fusion and decision-level fusion. In feature-level fusion, LiDAR point clouds and image features are combined or interacted with in a neural network [30]; in decision-level fusion, LiDAR and vision systems carry out separate reasoning, and their results are merged by voting or weighting [31]. The above methods have good results in the field of autonomous driving, but they all assume that LiDAR sensors can provide continuous, dense spatial point cloud data [32,33]. However, in rebar tying situations, single-point LiDAR sensors are generally used for construction cost reasons; these sensors produce binary signals and cannot be directly applied to the previous fusion framework. How to carry out effective multimodal detection under low-information LiDAR signal conditions is still a problem that current research has not yet solved completely.

2.3. Event-driven visual perception strategies

In recent years, with the increasing demand for vision systems, people have begun to pay attention to event-driven visual perception methods [34]. The main idea is that the vision module does not work all the time and will be started by an external signal (motion, sound, or distance trigger) to carry out

inference only when needed [35]. This strategy has already been used in intelligent surveillance and industrial inspection; however, there has been no systematic research on its application in rebar tying construction robots [36].

In the case of rebar tying, the trigger signal of the laser sensor can be used to start the event-driven method; A laser trigger indicates that there is something below the tying head, but it cannot determine whether it is a tieable node. By using the laser trigger as a visual activation signal and establishing a "trigger-verification" collaborative mechanism, we can maintain the fast response of the laser while introducing the precise discrimination ability of the visual system to achieve complementary advantages between the two modes.

2.4. Limitations of existing research and the scope of this work

In the previous section, three kinds of existing research have been introduced from all angles: system architecture, multimodal fusion methods, and event-driven perception. The current state-of-the-art in this area has the following problems.

(1) System Architecture: The current perception systems for rebar tying robots mainly use vision-based architectures. Vision sensors work all the time and have a high computing load, so they are not very stable in complex environments. There is no system architecture that can make multiple sensors work together at the task level yet.

(2) Multimodal Fusion Level: The existing fusion schemes use lasers to provide continuous point cloud data, but the actual hardware conditions in rebar tying scenarios are single-point lasers that output binary signals, so direct transfer and application are difficult.

(3) Event-Driven Perception Level: Event-driven visual activation strategies have not yet been systematically studied in the field of rebar tying construction robots, and there is a lack of collaborative mechanisms that combine trigger signals with visual verification.

To address the above problems, this paper puts forward a node-based perception system that can synchronise the firing of lasers and visual inspection. Unlike feature-level multimodal fusion, this paper introduces a "trigger-verification" task-level coordination mechanism at the system architecture level: binary laser signals are conditional activation signals for the visual module, and the visual module performs inference only after being triggered. Retaining the original laser hardware, this system has greatly improved the reliability of tying decisions by means of task-level coordination.

3. System overview and multimodal collaborative perception model

3.1. Operational scenarios and task definitions for reinforcement binding robots

Reinforcement tying is a critical process in the construction of prefabricated buildings and cast-in-place concrete structures. As shown in Figure 3a, the reinforcement tying robot moves above the reinforcement mesh and automatically ties the reinforcement intersections using its end-effector (tying head). The node detection range overlaps with the tying work area, ensuring that the recognition results can be directly applied to control the operation.

The rebar tying joints studied in this paper refer to structural locations where two or more rebars intersect or come into proximity in space, as shown in Figure 3. In images, these joints appear as areas

where rebar edges intersect; in laser trigger signals, they manifest as sudden changes in distance simultaneously detected by horizontal and vertical laser sensors.

Based on the above scenarios and definitions, this paper defines two core tasks for node recognition: node detection and node type recognition. The inputs, outputs, and decision-making processes for these two tasks are shown in Figure 4 and Table 1, respectively.

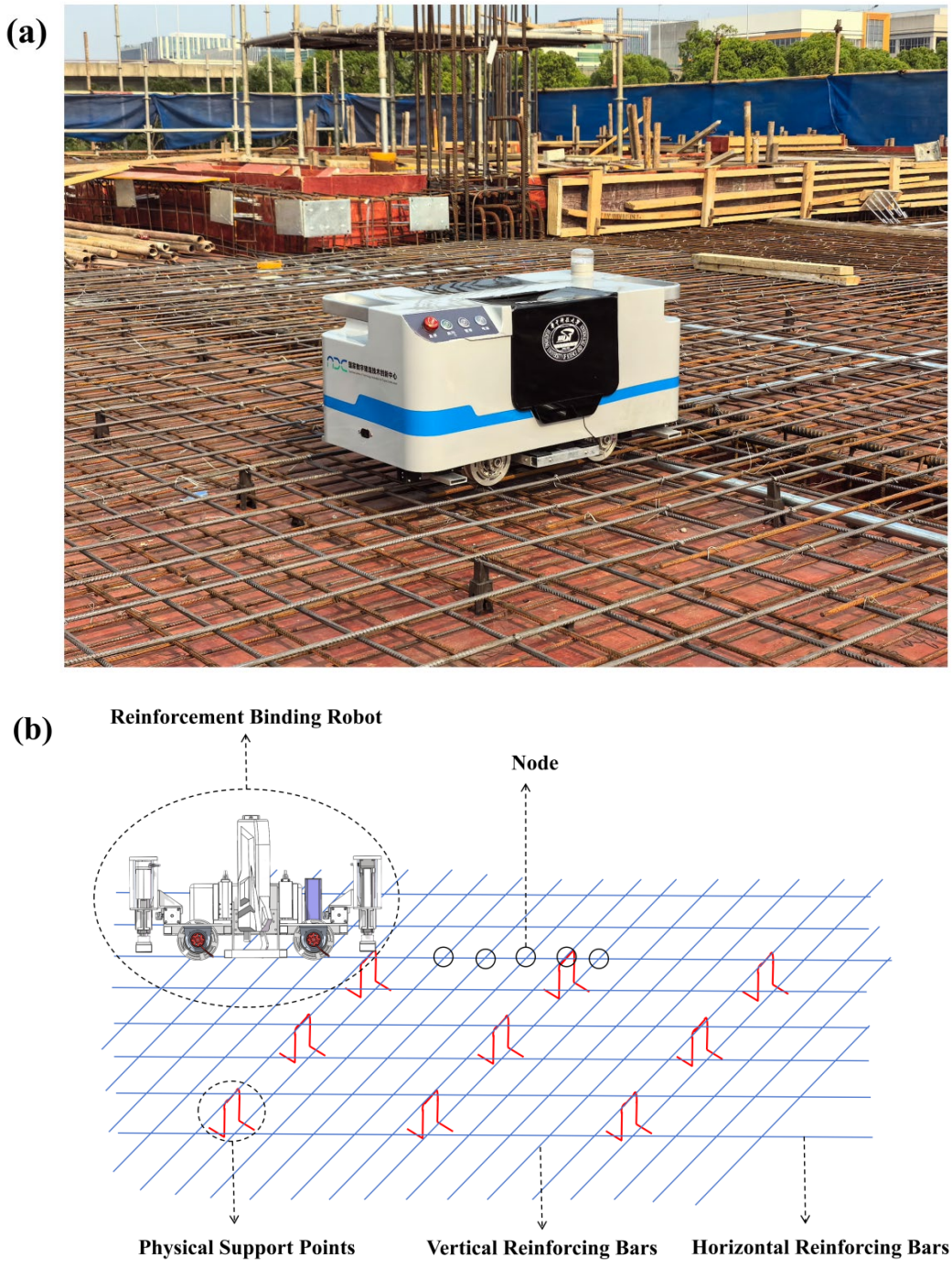


Figure 3. Operational scenes of the rebar tying robot. **(a)** Construction site view; **(b)** Schematic diagram of a joint.

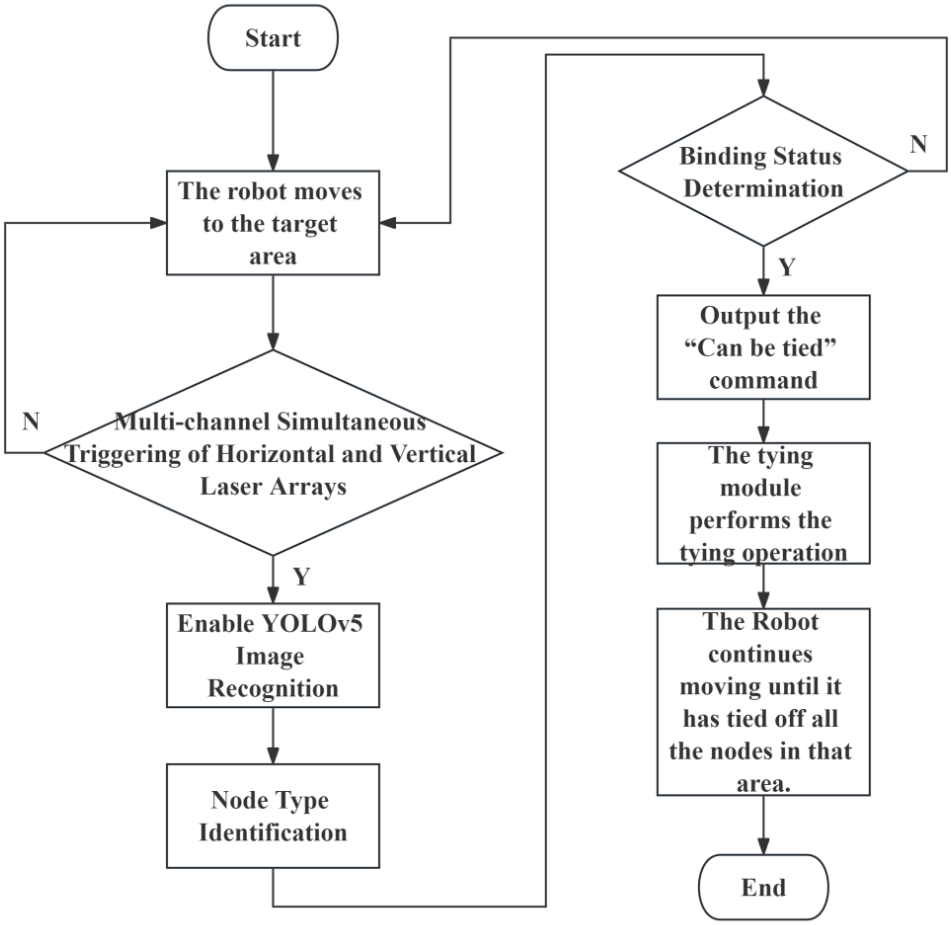


Figure 4. Flowchart of the node identification task.

Table 1. Node identification task definition table.

Task	Input Information	Output Format
Node Presence Detection	6 laser trigger signals (0/1)	Trigger flag + "Activate Vision Recognition" command
Node Type Identification	Local image of the node corresponding to the laser position	Node category + Strapping command

Node detection is performed using six laser trigger signals (0/1)—when both the horizontal and vertical laser groups are triggered simultaneously, the system determines that a node candidate exists within the current working area and generates an "Activate Visual Recognition" command. Node type recognition is performed by the YOLOv5 model after the laser is triggered. The model infers the node category from the local image captured beneath the tying head, providing a decision basis for the robot's subsequent tying actions.

3.2. Design of a multi-sensor collaborative perception system

To accomplish the node identification task described above, this paper proposes a bimodal collaborative perception system consisting of a visual camera and a single-point laser sensor. The two types of sensors

have clearly defined functional roles: the laser sensor provides highly reliable, low-information-content trigger signals used to determine the presence of nodes; the visual sensor provides high-information-content image data used to assess the node's environment and determine its binding status.

The sensor mounting positions are shown in Figure 5. The laser sensor layout is illustrated in Figure 5a, where six laser sensors (#1–#6) are mounted on both sides of the support plates and along the central axis of the perforated plate between the two idler wheels. Specifically: Laser sensors #1 and #2 are mounted on the left support plate at the centerline of the grooved pulley; #3 and #4 are mounted on the centerline between the perforated plate and the center of the binding head; and #5 and #6 are mounted on the right support plate at the centerline of the grooved pulley. The layout of the vision sensors is shown in Figure 5b, with the depth camera mounted above the binding head.

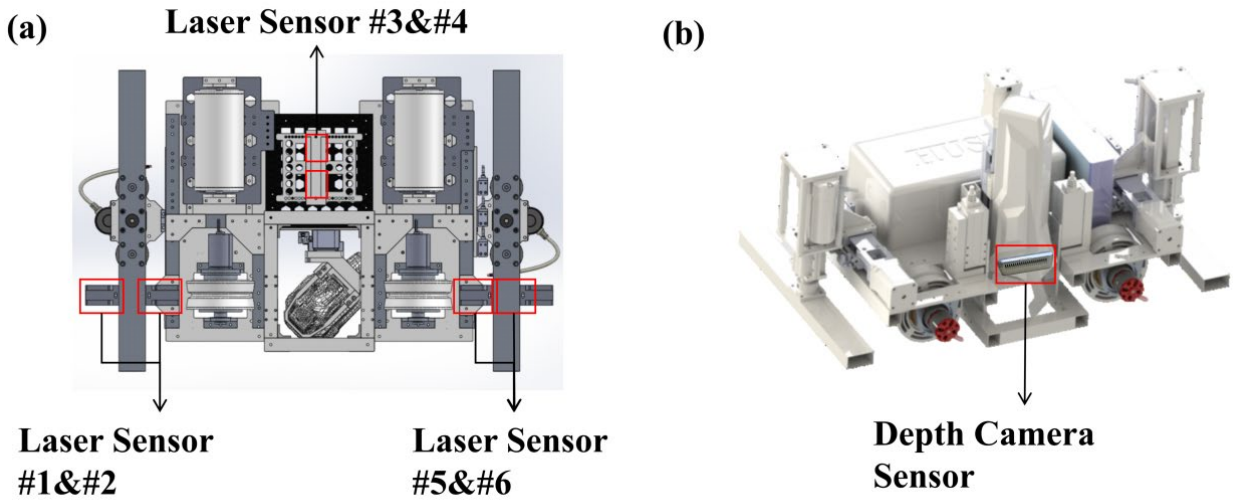


Figure 5. Schematic diagram of sensor layout.

The functional roles of the two types of sensors are shown in Table 2. The vision camera outputs RGB image frames, which are used to extract texture, edge, and intersection features of the nodes; the laser sensors output binary signals (0/1), indicating whether an object is detected within a preset distance range. Multiple laser sensors work in concert; when a preset combination simultaneously returns a value of 1, the system determines that a node candidate is present.

Table 2. Sensor function and location table.

Sensor	Mounting Location	Output Type	Functional Role
Camera	Above the tying head	RGB image frames	Node appearance extraction
Laser sensors (#1, #2, #5, #6)	Both sides of the support plate and the hollow plate	Binary signal (0/1)	Horizontally installed; detects transverse rebars
Laser sensors (#3, #4)	Center axis between the hollow plate and the tying head	Binary signal (0/1)	Vertically installed; detects longitudinal rebars

3.3. Laser triggering mechanism and vision activation strategy

This section explains how the laser sensors detect nodes through a "logical AND" triggering mechanism and how this mechanism activates the vision recognition module to achieve "trigger-verification" collaborative perception.

(1) Laser sensor layout

As shown in Figures 3 and 4, the system is configured with four horizontal laser beams and two vertical laser beams, whose detection areas do not interfere with one another. The horizontal laser beams are used to detect the presence of horizontal rebar, while the vertical laser beams are used to detect the presence of vertical rebar. As the robotic device moves along the horizontal rebar, a rebar intersection node is detected beneath the current tying head only when both the horizontal and vertical laser sets produce valid triggers. As shown in Figure 6a.

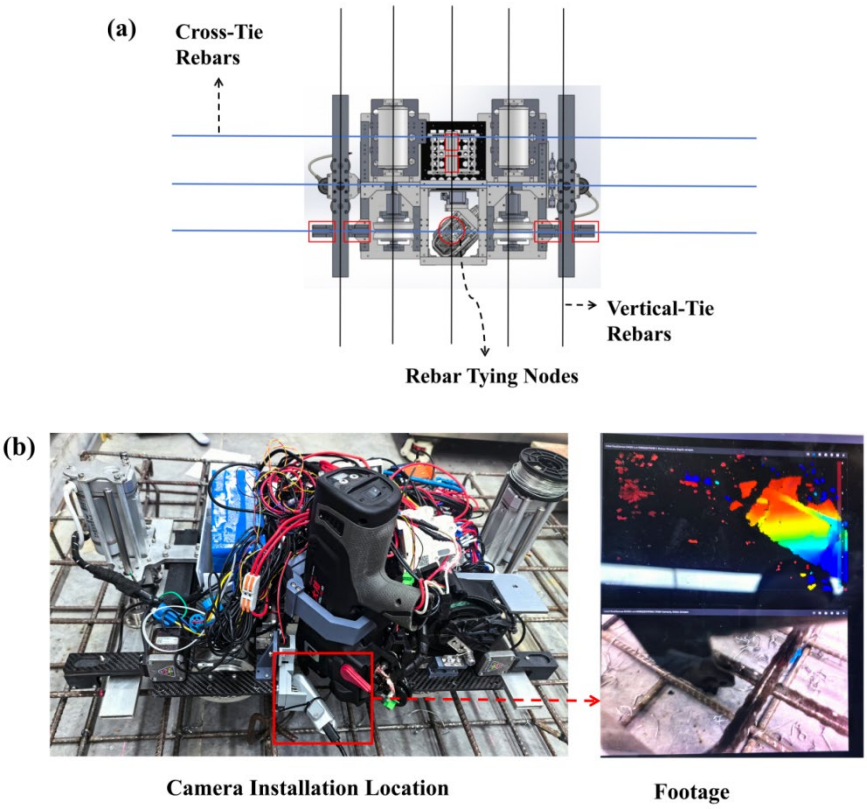


Figure 6. Schematic diagram of sensor layout and sensing areas. **(a)** Relative positions of the laser sensor and binding nodes; **(b)** Recognition area of the vision sensing module.

(2) Node presence detection

Let H denote the number of horizontal laser sensors that detect a target, and V denote the number of vertical laser sensors that detect a target. According to the hardware layout (Table 2), there are 4 horizontal sensors and 2 vertical sensors. The trigger condition for node detection is defined as:

$$S_{trigger} = \begin{cases} 1, & if\ H \geq 2\ and\ V \geq 1 \\ 0, & otherwise \end{cases} \quad (1)$$

When at least 2 of the 4 horizontal lasers return 1 (*i.e.*, $H \geq 2$), and at least 1 of the 2 vertical lasers returns 1 (*i.e.*, $V \geq 1$), $S_{trigger} = 1$. At this point, the system determines that a node candidate exists beneath the tying head and generates the "Activate Visual Recognition" command; otherwise, it is determined as "no node".

(3) Vision activation strategy

During robotic operations, the laser sensor continuously collects data at a high frequency, while the vision camera initially remains in a low-power standby state. When $S_{trigger} = 1$, the system activates the vision recognition module to capture and recognize images of the node area directly below the current tying head. After running once, the vision module returns to standby mode, awaiting the next laser trigger. This strategy focuses vision computing resources on actual node processing periods, avoiding the waste of computing power caused by continuous image inference.

The vision perception branch is the main validation unit in the multi-mode collaborative recognition system presented in this paper. After the laser trigger mechanism has determined whether there are nodes and where they are located, the visual perception module will perform image recognition on the area below the tying head to determine the exact type and status of the knot. Because the relative position of the tying head and the sensor is fixed, the visual module does not need to perform a global target search; it only needs to analyse the local image below the tying head, as shown in Figure 6, which can greatly reduce the complexity of visual recognition and inference latency.

In this paper, the main module for perception is YOLOv5. After the input images have been scaled to 640 × 640 pixels, they are then passed to the YOLOv5 network for feature extraction and object detection. The network outputs are object bounding boxes, class probabilities, and confidence scores; Figure 7 shows them.

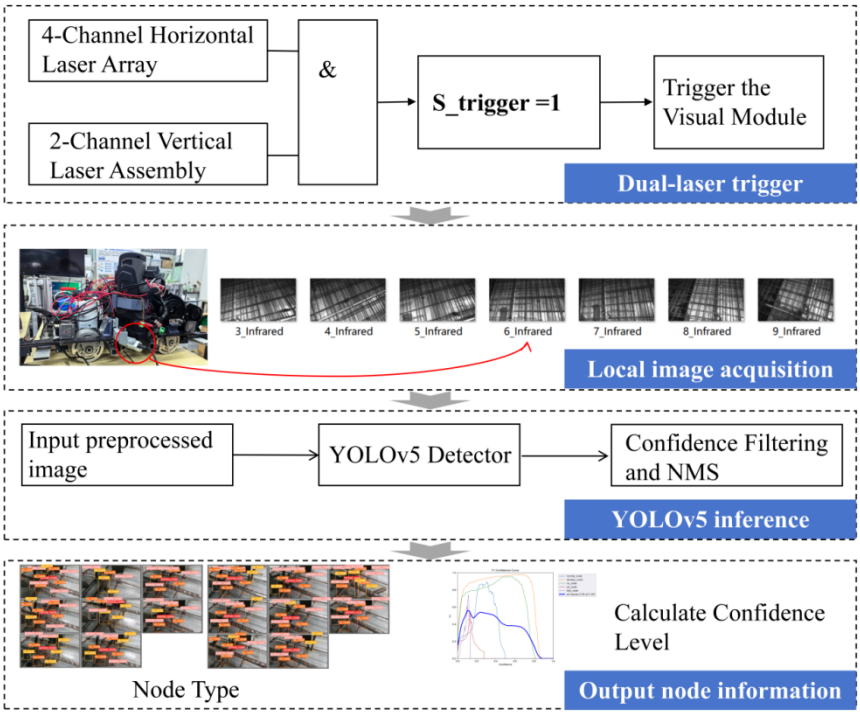


Figure 7. Flowchart of node image acquisition and recognition based on YOLOv5 under laser triggering.

YOLOv5 was trained with a supervised learning method and a set of annotated joint images during the model training phase. In the training stage, data augmentation techniques such as mosaic, random scaling, horizontal flipping, and Hue, Saturation, Value (HSV) colour space adjustment were adopted in order to increase the model's robustness to variations in lighting and differences in the surface of the reinforcement. After the training ends, the model can find the location of the joint in the input image and output its type label and tying state.

In the actual inference stage, the vision module will be turned on only if the laser trigger conditions are met. After both the horizontal and vertical laser groups are started, a local image is taken at the bottom of the tying head by the system, and then it is input into the trained YOLOv5 model for inference. The results of the inference include whether or not there is a joint and what type of joint it is.

(4) Necessity of using laser signals as the trigger

Although visual inference can also be initiated by the position of the robot, predefined waypoints, or other signals, these alternatives have significant engineering limitations in this tying scenario. First, since the robot moves straight along the rebar with grooved wheels, its position is inherently known and cannot provide physical information about whether there is actually a node underneath. Second, due to the dense distribution of nodes, a large number of coordinates would need to be set up manually, which would be very costly. Thirdly, from the results of the preliminary experiments, it can be seen that ultrasonic sensors are easily affected by reflections from rebar and dust, and thus the detection results are unstable. In contrast, a low-cost laser sensor mounted on the robot body can be used for real-time and accurate detection of the position of rebar intersections. Therefore, the best way to trigger the signal is to use a laser signal, as it is reliable, economical, and practical for engineering.

3.4. Multimodal collaboration mechanism

Different from the traditional way that sensors only serve as a camera trigger in a series of processes, the cooperative mode in this paper has reached the level of task division of labour; that is to say, the laser will confirm whether a node is physically present, and the vision module will determine its status. This division of labour is to free the vision module from the work of finding all the targets in the whole image; it only needs to classify the local image in the area that has been activated, so there is a good complementarity at the decision-making level between the two modalities.

3.4.1. Three-level decision logic

Once the vision module is activated, the following three-level sequence of decisions is executed:

Level 1: Node Confirmation. Determine whether the node actually exists based on the vision detection confidence S_κ. If $S_\kappa \geq T_{vision}$ (confidence threshold), the node is considered to exist; otherwise, it is considered to be “no node”, and a retry or skip will be initiated. It should be noted that because the camera is fixed at a certain height above the tying head, the image that is captured will be centred on the node directly under it. Therefore, there will be no more than one normal_node in the field of view corresponding to the current tying position. The system need not select from several candidate bounding boxes; if a normal_node with a confidence score greater than is detected, then a Tie command will be issued; otherwise, a Skip command will be issued.

Level 2: Node Category Determination. For nodes that have been confirmed to exist, determine whether the YOLOv5 output category of that node is a bindable node. If it is, go to Level 3; otherwise, output a "Skip" command.

Level 3: Binding Decision. For nodes determined to be valid, output a "Bindable" command. The three-level decision logic is shown in Table 3.

Table 3. Three-level decision logic table.

Level#1Node Confirmation	Level#2Category Determination	Level#3Binding Decisions
No nodes $S_k < T_{vision}$	/	Skip
There are nodes $S_k \geq T_{vision}$	Not tied down / No abnormalities	Output the "Can be tied" command
There are nodes $S_k \geq T_{vision}$	Binding/Offset/Crossing Errors	Skip and report to the backend

3.4.2. Confidence-based exception handling

When the vision system is unable to detect a node more than once in succession, the system will take exception-handling measures as follows:

(1) Retry Mechanism: The system will keep track of the number of consecutive failures. If the number is less than or equal to the maximum retry limit (3 times), then the robot will make some small posture adjustments and re-trigger the perception process.

(2) Adaptive adjustment: If there are too many consecutive failures, the system will reduce the laser trigger conditions to improve the sensitivity of the trigger.

(3) If the attempt is still unsuccessful after modification, the current area will be marked as an "Environmental Anomaly" or "Node Anomaly" and reported to the host computer; the robot will ask for manual inspection and stop operation for maintenance.

3.5. Task output and post-processing interface

Once multimodal collaborative decision-making is complete, the system will encapsulate the results in the form of standardised commands and send them to the robot's execution unit through the control interface. The main task of the post-processing phase is to convert the output results of the three-level decision-making logic ("Tie" or "Skip") into action commands that can be executed by the controller. When the decision result is "Tie", the system generates a trigger signal and sends it to the underlying controller to drive the robot to execute the tying action; when the decision result is "Skip", the system outputs a skip command, and the robot continues to move to the next target position. This command-mapping mechanism lays the foundation for future deep integration of the vision algorithm and low-level motion control.

4. Dataset construction and experimental design

4.1. Experimental platform and hardware configuration

This section describes the hardware platform and software environment used in the experiments described in this paper. The experimental platform consists of the rebar tying robot itself, a multi-sensor perception unit, and a computing unit.

(1) Robot platform

The robot itself is an in-house developed experimental platform for rebar tying robots, as shown in Figure 8. This platform employs a four-wheeled mobile chassis equipped with a tying gun, enabling movement in the X, Y, and Z directions above the rebar mesh. The working area at the tip of the tying head measures 200 mm × 150 mm, with a maximum tying depth of 100 mm.

(2) Sensor configuration

The sensor system consists of a vision camera and a laser sensor, with their mounting positions shown in Figure 8d. The vision camera is an Intel RealSense Depth Camera D435i, with an RGB resolution of 1920 × 1080 pixels, a frame rate of 30 frames per second, and a calibrated color camera lens focal length of 6 mm. The camera is mounted on the horizontal side of the tying head and angled downward toward the work area at a 40 ° angle. The laser sensors consist of six DS-GLS31 photoelectric switch-type spot/line laser sensors, as shown in Figure 8b and c. The detection range of each laser sensor is 20–300 mm, and a binary signal (0/1) is output; that is, a 1 means an object has been detected at a distance less than the set threshold, and a 0 means no object has been detected.

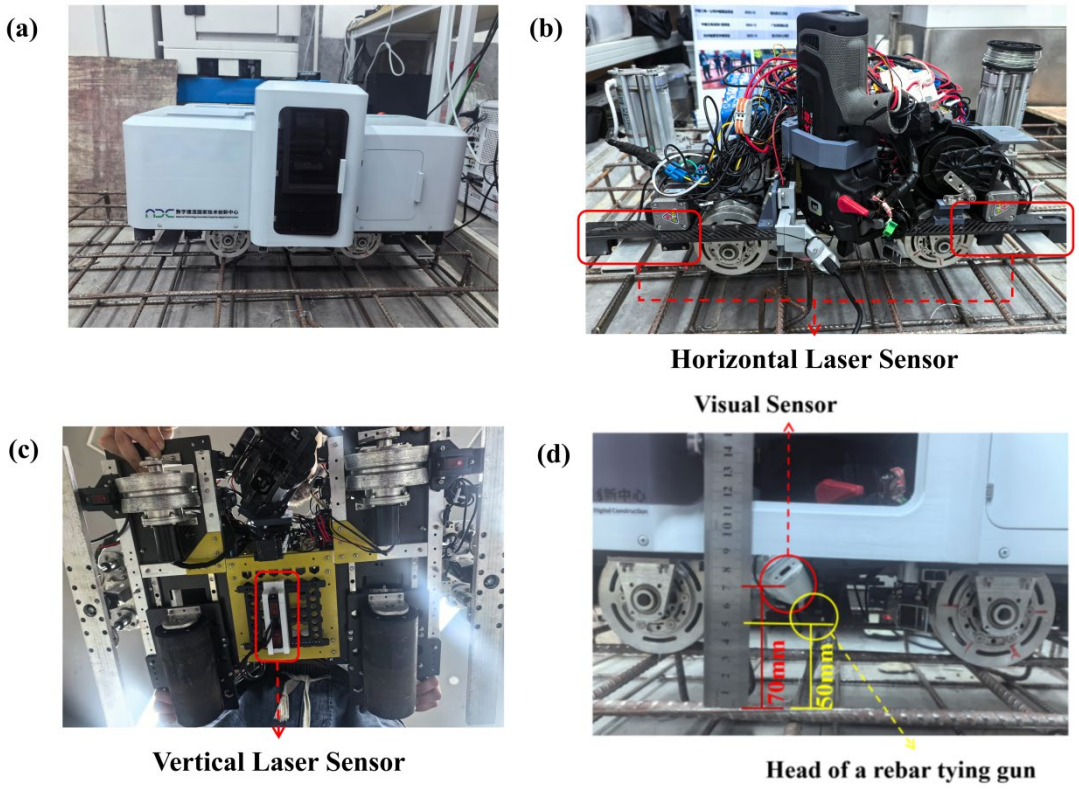


Figure 8. Schematic diagram of the robot chassis and sensor mounting locations.

(3) Computing unit and software environment

The robot control system is a distributed hardware system; the main computation and scheduling unit is the Zmotion motion controller. The controller communicates with ultrasonic sensors, encoders, and other peripherals via the RS485 serial bus; it also receives feedback signals from proximity switches, diffuse reflection sensors, and electric actuators via a digital input interface; then it drives the stepper

motor and binding gun to perform the binding operation via a pulse and digital output interface. Because there is no vision interface on the current Zmotion controller, a visual perception module has been added separately through a reserved interface. Once the laser trigger condition is met, the captured node images are transmitted to a host PC equipped with an Intel Core i7-12700H processor for offline inference using the PyTorch framework, with an average inference time of approximately 50.2 ms per frame. The entire system is powered by a lithium battery, which provides stable voltage to the controller, sensors, and actuators through a voltage regulator. Future work will investigate deploying lightweight vision models directly on the low-level controller.

4.2. Multimodal data collection and annotation methods

(1) Data collection scenarios

In order to prove the feasibility of the method proposed in this paper, some multimodal data were collected in daily life rebar-tying situations. Data were collected in both indoor laboratory and outdoor construction site environments with the same lighting conditions (normal illumination of 500–1000 lux), different degrees of occlusion (no occlusion, partial occlusion), and the same rebar specifications (10 mm diameter). The rebar mesh was laid with horizontal and vertical rebars intersecting, with a grid spacing of 150 mm × 150 mm.

(2) Data collection process

The data collection process is as follows: The robot moves along a preset path, with the tying head moving at a constant speed at a fixed height above the rebar mesh. The laser sensors continuously collect data at a frequency of 50 Hz, while the vision camera initially remains in standby mode. When the laser trigger conditions are met (at least 2 of the 4 transverse rebar laser sensors simultaneously return a value of 1, and at least 1 of the 2 vertical rebar laser sensors returns a value of 1), the vision sensor is activated. The system will store the current image frame and the corresponding laser signal, and it will also keep a timestamp. 300 valid frames have been taken, and 2660 rebar joints are included in them. Table 4 shows the results.

(3) Data annotation method

Data annotation was performed using the LabelImg image annotation tool. As shown in Figure 9, the annotation content and rules are as follows:

(a) Joint Bounding Box: Mark the smallest rectangle that contains the rebar intersection joint; this is based on the coordinates of the four corners of the rectangle.

(b) Node Status Categories: Classify them into five categories: normal_node (unbound and no anomaly), tied_node (bound), ob_node (obstructed by upper-layer rebar or debris), already_node (nodes already fixed in other layers), and no_node (crossings visible in the image but not actual nodes).

(c) Use the centre of the rebar intersection as the centre of the bounding box; one box will be created for each node. Each box should cover only the node and a little bit of surrounding rebar, and the overlapping areas should be reduced as much as possible.

Table 4. Statistical summary of the dataset.

Data Type	Count	Description
Total Image Frames	300	Includes triggered and non-triggered frames
Number of Node Annotations	2660	Reinforcement Intersection Nodes
normal_node	148	Normal condition; can be tied
tied_node	125	Tying completed
ob_node	256	Obstructed by debris; physical support interference present in the surrounding area
already_node	1076	Nodes already fixed on other layers
no_node	1055	Not real physical nodes, but visual misalignments in the 2D image.

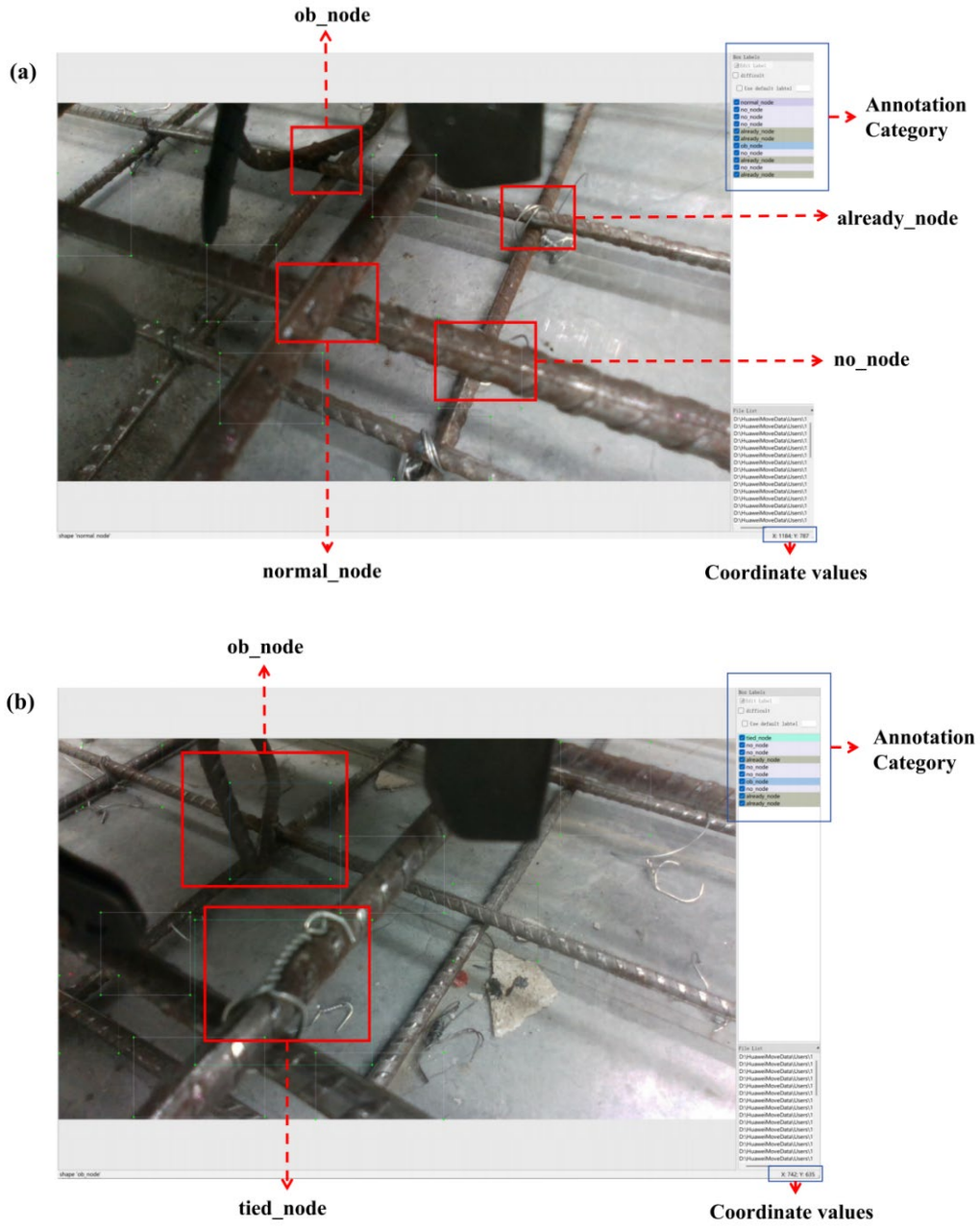


Figure 9. Data collection scenarios and annotation examples.

4.3. Data preprocessing and sample partitioning

The original images were obtained by the industrial camera at a resolution of 1920 × 1080 pixels. To meet the input requirements of the YOLOv5 network, all images were resized to 640 × 640 pixels and normalised to the range [0,1]. The specific training configuration is shown in Table 5: the YOLOv5s variant was trained on a CPU for 200 epochs with a batch size of 8. The Stochastic Gradient Descent (SGD) optimizer was employed with initial and final learning rates of 0.01, a momentum of 0.937, and a weight decay of 0.0005. During training, data augmentation strategies including Mosaic, random horizontal flipping (probability 0.5), and HSV color space adjustment were applied. During the inference phase, the Non-Maximum Suppression (NMS) threshold was set to 0.45; the confidence threshold T_{vision} for the Tie/Skip decision was set to 0.5, which was selected based on the Precision-Recall (PR) curve on the validation set to balance the false-binding rate and missed-tie rate.

After preprocessing, the 300-image dataset was randomly split at the image level into training (60%, 180 images), validation (20%, 60 images), and test (20%, 60 images) sets, as shown in Table 6. The training set was used for model parameter learning, the validation set for monitoring overfitting, and the test set solely for final performance evaluation.

Table 5. Training configuration and hyperparameters.

Parameter	Value
Model variant	YOLOv5s
Input size	640 × 640
Batch size	8
Epochs	200
Device	CPU
Optimizer	SGD
Initial / Final Learning Rate (LR)	0.01/0.01
Momentum	0.937
Weight decay	0.0005
Anchors	Auto-computed
NMS threshold	0.45
Confidence threshold (T_{vision})	0.5

Table 6. Sample classification table.

Dataset	Number of Images	Number of Nodes	normal_node	already_node	no_node	ob_node	tied_node
Training Set	180	1608	90	658	632	154	74
Test Set	60	528	29	218	216	38	27
Validation Set	60	524	29	200	207	64	24
Total	300	2660	148	1076	1055	256	125

4.4. Experimental design

To verify the effectiveness of the "laser triggering + visual verification" collaborative mechanism proposed in this paper, we designed comparative experiments to evaluate the improvement in system performance resulting from the addition of a visual module within the same physical scenario.

(1) Experimental scenario and testing methods

The experiments were conducted on a completed rebar mesh. This scenario included multiple rebar intersection nodes, covering various node states (tied, untied, abnormal, *etc.*). The robot traversed each node sequentially along a preset path. Within the same scenario, two perception schemes were run separately, and their respective recognition and execution results were recorded. The images used for the experimental comparison were independently collected, and all were captured by the camera triggered at the moment the laser sensors confirmed the presence of a node.

(2) No-vision scheme (S1: laser trigger only)

This approach employs triggering logic based on four horizontal and two vertical laser beams. When both the horizontal and vertical laser groups are triggered, the system determines that a tieable node exists directly beneath the tying head and immediately outputs a "tieable" command. This approach does not include any image recognition or state determination steps.

(3) Vision-based approach (S2: laser triggering + YOLOv5 verification)

After the laser triggering conditions are met, the system captures a local image beneath the tying head and performs inference using the YOLOv5 model. The system will only issue a "Tieable" command if the model's detection result is "normal_node" and the detection confidence is above the set threshold. Otherwise, it will send a "Skip" or "Anomaly" command. The experimental setup is shown in Table 7.

Table 7. Comparison of experimental setups.

Solution ID	Solution Description	Decision Criteria	Test Scenario
1	Laser detection only	A laser trigger is automatically classified as a normal_node	#1 Reinforcing Mesh, Same Path
S2	Laser trigger	Laser Trigger + YOLOv5 to determine if it is a normal_node	#1 Reinforcing Mesh, Same Path
S3	Vision-only continuous inference	Continuous YOLOv5 inference confirms normal_node	Rebar Mesh #1, same path

4.5. Evaluation metrics

The performance of the system will be evaluated in this paper using the following four indicators: precision, recall, F1 score, and false-binding rate (FBR).

Precision measures the proportion of true nodes in the detection results:

$$P=\frac{TP}{TP+FP} \quad (2)$$

Recall measures the proportion of true positive cases detected:

$$R=\frac{TP}{TP+FN} \quad (3)$$

The F1 score is the harmonic mean of precision and recall, providing a comprehensive measure of detection performance:

$$F1=2\times\frac{P\times R}{P+R} \quad (4)$$

The false-binding rate (FBR) is defined as:

$$FBR=\frac{FP}{TP+FP}=1-\text{Precision} \quad (5)$$

TP is the number of correctly identified tieable nodes, and FP is the number of non-tieable nodes that were incorrectly identified as tieable. Though they are mathematically the same, Precision and the false-binding rate represent system performance from different angles; the former focuses on the proportion of correct binding decisions, and the latter directly measures the safety risk of incorrect binding.

It is worth noting that in the case of rebar tying, not all false bindings will result in problems with structural safety. If the system misidentifies a tied_node as a tieable node, it only causes redundant tying operations, which mainly increase wire consumption and do not have any negative effects on the mechanical properties of the structure. Conversely, if a true normal_node is missing, then the structural connection strength will decrease. Therefore, the following analysis will focus more on the system's rejection ability for ob_node and the missed-tie rate.

5. Results and discussion

5.1. Visual module detection performance

5.1.1. Overall detection mAP@0.5 and performance by category

The detection performance (mAP@0.5) of the YOLOv5 model was evaluated on a 10 mm rebar node dataset (300 images containing 2660 annotated nodes). Using the training parameters described in Section 4, the model achieved an overall mAP@0.5 of 92.0% on the 60-image held-out test set. The detection results for each category are presented in Table 8.

Table 8. Detection performance (mAP@0.5) by category on the test set.

Node Type	Number of Images	Number of Instances	P	R	mAP@0.5
normal_node	60	29	0.966	1.000	0.968
already_node	60	218	0.924	0.895	0.945
no_node	60	216	0.947	0.926	0.968
ob_node	60	38	0.790	0.658	0.736
tied_node	60	27	0.993	0.963	0.984
Total	60	528	0.924	0.888	0.920

As can be seen from Table 8:

The model demonstrates excellent detection performance for tieable nodes (normal_node), achieving an mAP@0.5 of 96.8% and a Recall (R) of 1.000. This indicates that the model can identify 100% of tieable nodes in practical operations, thereby avoiding the risk of missed ties. Furthermore, the tied nodes (tied_node) achieve an mAP@0.5 of 98.4% and a Precision (P) of 99.3%.

For common pseudo-nodes (no_node) and already fixed nodes (already_node) in the images, the model achieves mAP@0.5 values of 96.8% and 94.5%, respectively, effectively distinguishing them from nodes to be tied.

For obstructed or physically interfered nodes (ob_node), the mAP@0.5 reaches 73.6%. Although the performance is relatively lower than other categories due to the lack of visual features caused by occlusion, it has improved significantly compared to previous experiments and meets engineering requirements.

5.1.2. Analysis of convergence in the training process

As can be seen from Figure 10a, the model has good performance on the test set. It can correctly identify and locate various nodes, such as normal_node, already_node, and no_node, and is quite stable in detecting dense, complex, and partially occluded areas. Green boxes are the predicted results, and red boxes are the ground truth. It can be seen that for the majority of nodes (especially the important normal_node and tied_node), the predicted boxes are in good agreement with the ground truth, and there are few false or missed detections.

Figure 10b is a diagram of the loss and accuracy curves for the 200-epoch training process. From the figure above, it can be seen that the three loss functions are all decreasing rapidly in the initial stage and tend to stabilise about 50 epochs later. The validation loss (val_loss) also converges steadily without many changes; therefore, the model is well-trained, and there is no overfitting in box_loss, obj_loss, or cls_loss. At the same time, Precision, Recall, and mAP@0.5 steadily increase with the training epochs and finally reach above 0.9 in mAP@0.5 s, which is good for the model.

From the Precision-Recall (PR) curves of the test set in Figure 10c, it can be seen that the model has an excellent recognition effect for all node categories, and its mAP@0.5 is very high. The two curves for tied_node and normal_node are closest to the top-right corner, with their respective mAP@0.5 values being 98.4% and 96.8%. The recall for normal_node is 1.0, and it can identify all the tieable nodes in the system. The mAP@0.5 of ob_node (obstructed nodes) is slightly lower than that of other categories because some visual features are missing; however, the general shape of the curve is still good. While most categories achieve high mAP@0.5 values above 92.0%, the ob_node class achieves a relatively lower mAP@0.5 of 73.6% due to severe occlusion. Therefore, the overall performance demonstrates that the YOLOv5 model is suitable for node recognition tasks.

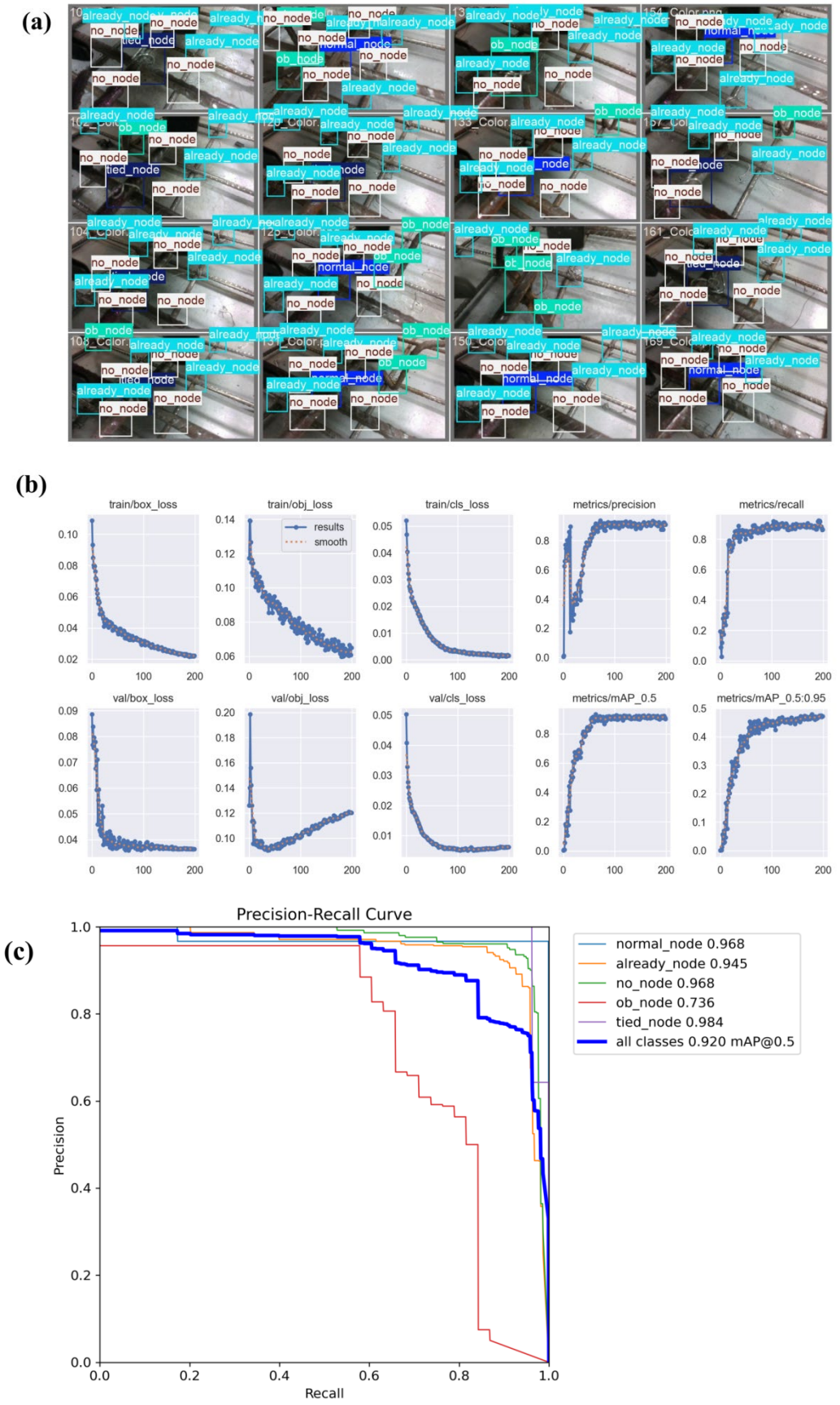


Figure 10. Visualization of the YOLOv5 model's detection performance and results. **(a)** Example of detection results on the validation set; **(b)** Loss and accuracy curves during training; **(c)** Precision-Recall curves for the test set.

5.2. Comparison of visual and non-visual approaches

To verify the performance of the vision module in actual joint recognition, 91 images of rebar joints were selected as test samples; that is, the inference and decision-making processes for the three methods, S1 (laser-only), S2 (laser-triggered + YOLOv5 verification), and S3 (vision-only continuous inference baseline), were simulated. We have checked the 91 images at both the image and acquisition-sequence levels to make sure that they were not used in the training, validation, or testing of the model; thus, the data for comparison is independent.

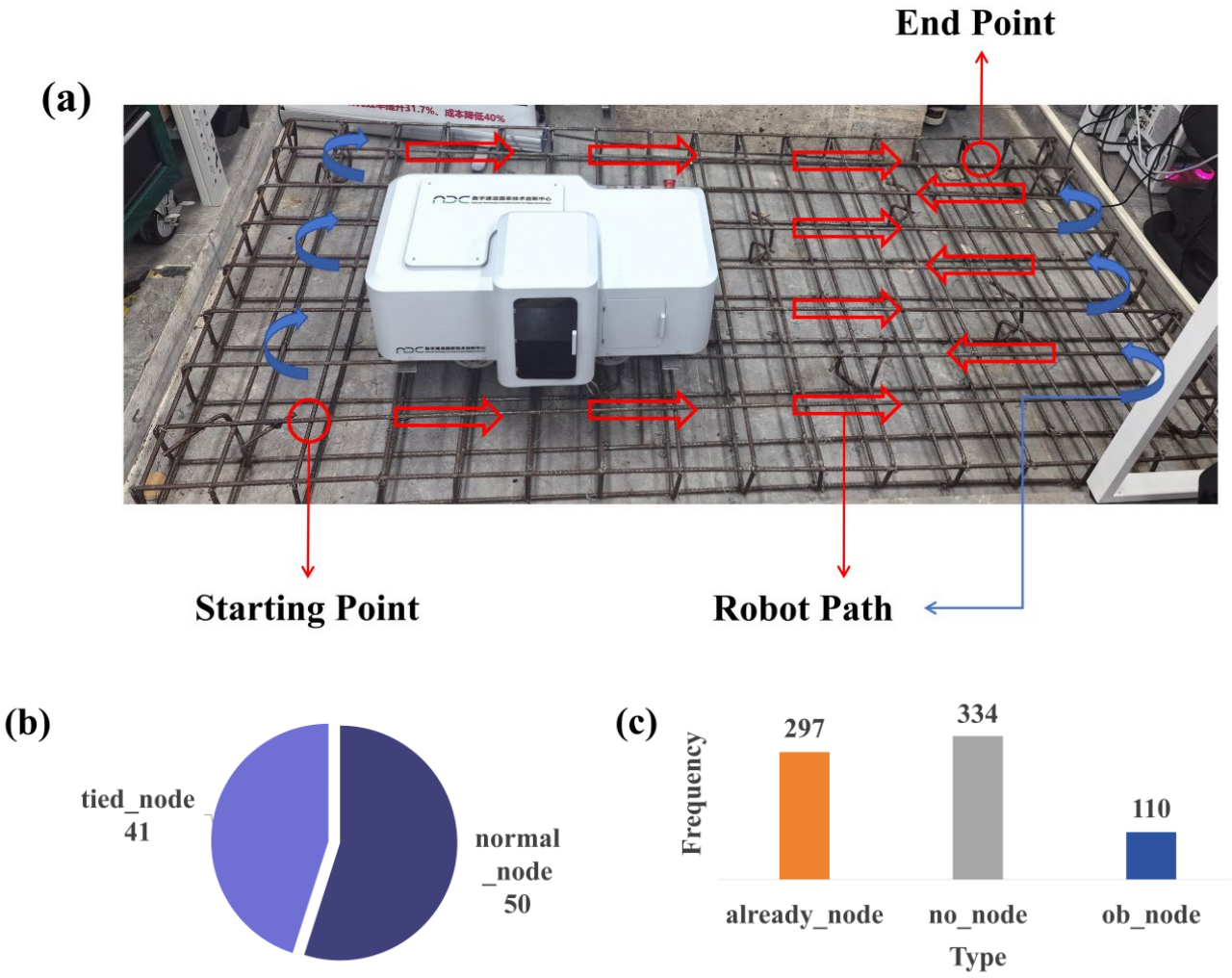


Figure 11. Experimental setup and distribution of test samples. **(a)** Schematic diagram of the robot's path; **(b)** Distribution of physical node categories; **(c)** Frequency of background interference.

The experiment was carried out on Rebar Mesh #1, where the robot went through each node in turn along a predetermined path as shown in Figure 11a. A total of 91 test images were obtained. There is only one target node under the tying head in every picture; this target node has been confirmed by the laser trigger. The category distribution is shown in Figure 11b, which contains 50 normal_node (tieable) and 41 tied_node (non-tieable). Besides, every image has some background interference elements such as already_node, no_node, and ob_node, whose statistics are shown in Figure 11c. The interference elements do not count as target nodes individually; however, they make the visual recognition task more complex.

Table 9. Summary of comparison results between S1 and S2.

Plan	Precision (P)	Recall (R)	False Binding Rate	Inference frames	Total inference time (ms)
S1	54.9%	100%	45.1%	0	0
S2	91.8%	100%*	8.2%	45	2259
S3	91.8%	100%*	8.2%	91	4568

* The recall of S2 and S3 is marked as 100% based on the strict test of 91 physical nodes (45 correct ties, 0 missed ties).

Analysis of Table 9 reveals that:

There has been a considerable improvement in decision accuracy; the precision of S2 is 36.9 percentage points higher than that of S1, and it can eliminate many false alarms caused by laser triggering through the visual verification module. Both S2 and the vision-only baseline (S3) have a false-binding rate of 8.2%; thus, it can be concluded that adding vision is one way to improve the reliability of multimodal decision-making.

Significant optimization in computational efficiency: In the test sequence containing 91 nodes, the vision-only baseline (S3) requires continuous inference on all 91 frames, whereas the proposed method (S2) relies on laser triggering and performs visual inference on only 45 frames (a trigger rate of 49.5%). On the host PC equipped with an Intel Core i7-12700H processor, the average inference time per frame is 50.2 ms. Based on this, S2 reduces the YOLO inference workload by approximately 50.5% compared to S3. It should be noted that this reduction specifically refers to the number of inference operations and does not include image acquisition or communication overhead. The specific decision output and time statistics interface of the system running on the host PC are shown in Figure 12.

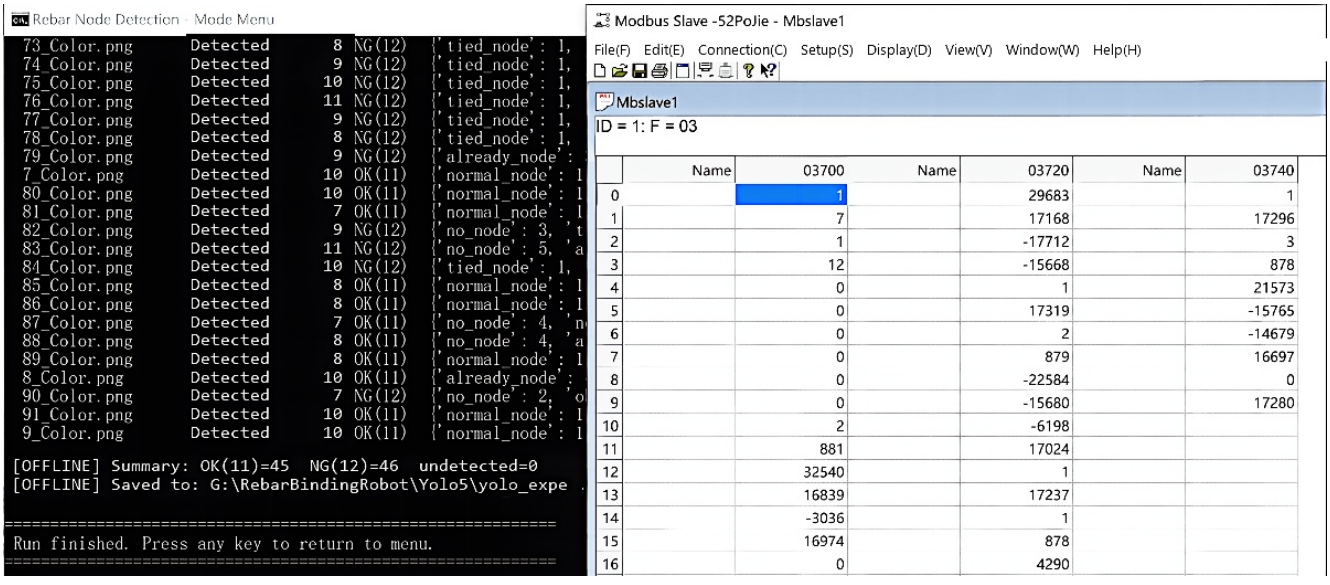


Figure 12. Visual decision outputs and Modbus register status of the host-controller communication.

Nature of false bindings and safety impact: The 8.2% false-binding rate of S2 in the macro test mainly occurred when tied_node was misclassified as a tieable node. As discussed in Section 4.5, this type of false binding only increases material consumption and does not affect structural safety. In the test on the 91 image frames (see Table 10), the system achieved 0 false bindings and 0 missed bindings,

accurately filtering out all anomaly interference items such as ob_node, which proves the high reliability of the visual decision mechanism in terms of core safety.

Table 10. Three-way confusion table of S2 decisions on the 91 test images.

Decision Category	Count	Percentage
Correct tie (Tieable node,output OK)	45	49.5%
Correct skip (Non-tieable node,output NG)	46	50.5%
False tie (Non-tieable node,incorrectly output OK)	0	0.0%
Missed tie (Tieable node,incorrectly output NG)	0	0.0%
Total	91	100%

System boundary between perception and execution: In real robotic operations, if the vision system correctly identifies a normal_node but the tying gun fails to complete the tying due to mechanical interference (e.g., jamming or slipping), this belongs to the reliability issue of the execution mechanism, rather than a failure of the visual perception decision. This paper focuses on validating the effectiveness of the multimodal perception and decision-making layer. Hard real-time closed-loop validation at the execution level will be carried out as future work.

Robustness and typical faults are: Because the robot is driven by grooved wheels to roll along the rebars, if there is lateral movement of the robot, the wheels will slip and raise an alarm. Therefore, lateral deviation will not occur during normal operation. If highly irregular rebar placement causes the physical position of a node to move beyond the mechanical tolerance of the fixed tying gun, the system will stop the tying action and report it for manual handling, thus preventing blind operation.

6. Conclusions and future work

6.1. Research conclusions

In this paper, we address the problem of joint detection for rebar tying robots in complicated construction environments and put forward a multimodal perception solution that uses laser triggering and visual verification. Based on YOLOv5, this study developed a complete technical framework for data collection and system deployment. According to the above comparative experiment and efficiency analysis, we can get the following conclusions:

(1) The YOLOv5-based visual recognition module can detect 10 mm rebar nodes well. The overall mAP@0.5 of the five node categories in the validation set is 92.0%. In particular, the detection performance values for tied_node and already_node are 98.4% and 94.5% respectively, respectively, and normal_node achieves 96.8%; thus, it is suitable to apply the YOLOv5 model in node recognition tasks.

(2) Compared with a purely laser-based solution, adding visual verification reduced the false binding rate of the system from 45.1% to 8.2%, and the accuracy rate increased from 54. The results show that the visual module can distinguish well between bindable and non-bindable joints; therefore, it is more reliable and safer in binding operations 9%, 91.8%.

(3) Compared to the vision-only continuous inference scheme (S3), the laser-triggering mechanism in our method makes the vision module process only 45 frames (a trigger rate of 49.5%), reducing the

YOLO inference workload by about 50%. It should be noted that this trigger rate will vary dynamically according to the node density in the operating environment by 5%.

The above results validate the feasibility and engineering potential of the proposed "laser-triggering + visual-verification" collaborative mechanism in node perception tasks for rebar tying robots. In addition, the method is also well-suited for cost-sensitive construction projects in terms of hardware costs. Conventional vision-only architectures usually use high-resolution cameras and powerful computing units, and multi-sensor fusion architectures require expensive LiDAR, while our method only uses extremely low-cost spot/line laser sensors and a standard depth camera. By maintaining high accuracy (92.0% mAP@0.5), this cost-effective hardware combination significantly lowers the procurement threshold and is highly suitable for cost-sensitive engineering demands.

6.2. Limitations and future work

Although the proposed method has achieved promising results in experiments, limitations remain in terms of cross-scenario generalization and system-level integration. To address these deficiencies, future research will proceed in the following two aspects:

(1) Expansion of cross-scenario generalization and complex 3D structures. Although the current dataset has been expanded to 300 images, all data were collected from the same horizontal rebar mesh, and the image-level random split may lead to correlated visual features across subsets. This limits the evaluation of cross-mesh generalization. Furthermore, there is still an issue of uneven class distribution. Future work will focus on supplementing node images under different rebar specifications, multiple lighting conditions, and severe occlusion environments, and exploring synthetic data techniques to enhance model robustness. Meanwhile, for complex scenarios such as multilayer rebar meshes or 3D rebar cages, we plan to investigate adaptively adjustable laser-triggering arrays and introduce PTZ (pan-tilt-zoom) cameras with adjustable viewpoints and tying modules with multiple degrees of freedom to accommodate rebar nodes of different heights and orientations.

(2) Optimization of system closed-loop integration and physical alignment. Constrained by the current controller interface, vision inference is performed on a host PC, and complete end-to-end latency testing (including communication and execution) has not yet been conducted. Additionally, because the tying gun is fixed relative to the robot body, there is a potential risk of physical misalignment when the on-site rebar placement deviation is too large (currently, a "halt-and-report" mechanism is used for safety fallback). In the future, the perception and decision-making module will be deeply integrated into the robot's low-level control system to conduct full closed-loop tests encompassing triggering, communication, mechanical alignment, and tying actions. Furthermore, mechanical dynamic adjustment will be incorporated to optimize physical alignment accuracy, and the system will be compared with manual tying quality under real working conditions to provide direct data support for engineering acceptance.

Data availability statement

The data and datasets generated or analyzed in this study are included in the article and the supplementary material. The specific training code, dataset configuration file (data.yaml), hyperparameter file (hyp.yaml), training options file (opt.yaml), and dataset split script are provided as supplementary files.

Declaration of generative AI and AI-assisted technologies

During the preparation of this manuscript, the authors used generative AI tools (DeepSeek) only to improve language and readability. The authors reviewed and edited the content as needed and take full responsibility for the content of the manuscript.

Acknowledgments

The authors would like to thank National Key Research and Development Program of China (2023YFC3806903) for financial support.

Authors' contribution

Conceptualization, J.L. and Y.Z.; methodology, J.L.; software, J.L.; validation, J.L., Y.H. and X.Z.; formal analysis, J.L.; investigation, J.L. and S.W.; resources, Y.Z.; data curation, J.L.; writing—original draft preparation, J.L.; writing—review and editing, J.L., Y.Z. and C.Z.; visualization, J.L.; supervision, Y.Z.; project administration, Y.Z.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Conflicts of interest

Cheng Zhou holds the position of Associate Editor for *Smart Construction* and has not peer reviewed or made any editorial decisions for this paper. The authors declare no other conflicts of interest.

References

[1] Chen J, Zhao Z, Huang G. Localization and navigation of construction robots: a human-centric review. *Autom. Constr.* 2026, 187:106953.

[2] Wang T, Mao C, Sun B, Li Z. Genealogy of construction robotics. *Autom. Constr.* 2024, 166:105607.

[3] Jang H, Yi Y, Yu Y, Saqan EI, Bayrak O. Shear behavior of prestressed box beams with high-strength shear reinforcement. *Eng. Struct.* 2026, 351:122049.

[4] Tuncer E, Karageyik C, Canbay E, Muka N, Binici B. Seismic performance of post-tensioned cast-in-place reinforced concrete structures based on component testing. *J. Build. Eng.* 2026, 122:115744.

[5] Heydari M. Automation of rebar tying in the precast concrete manufacturing process. Master's Thesis, University of Georgia, 2021.

[6] Song Z. Population aging and its impact on China's high-quality economic development. In *Proceedings of the 2025 5th International Conference on Enterprise Management and Economic Development (ICEMED 2025)*, Dali, China, May 30–June 1, 2025, pp. 992–998.

[7] Akanmu A, Abraham Y. Neurosymbolic robotics for industrialized construction: a review of human-robot collaboration and explainable automation. *Adv. Eng. Inform.* 2026, 76:105074.

[8] Wang H, Jin Z, Shao Y. Geometry-driven perception and action for topology-agnostic robotic rebar tying. *Autom. Constr.* 2026, 188:107013.

[9] Guo J, Lyu X, Deng L, Xu S, Liu M, *et al.* Integrated 3D coverage path planning for robotic rebar tying in prefabricated construction. *Adv. Eng. Inform.* 2026, 76:105076.

[10] Hou S, Wu M, Wu B, Ding Y, Zhang J, *et al.* Operation path planning for rail-type rebar binding robots based on improved ant colony algorithm. *J. Comput. Civ. Eng.* 2025, 39:04025062.

[11] Liu M, Guo J, Deng L, Lyu X, Nie L. Advanced vision-guided robotic tying with 6-DoF rebar pose estimation. In *Proceedings of the 42nd International Symposium on Automation and Robotics in Construction*, Montreal, Canada, July 28–31, 2025, pp. 202–209.

[12] Dong L, Zhang R, Qin H, Shao Z, Wang X. Geometric error modeling and compensation for a tracked autonomous rebar binding robot. *Ind. Robot* 2026, 53(6):1283–1291.

[13] Cheng B, Deng L. Vision detection and path planning of mobile robots for rebar binding. *J. Field Rob.* 2024, 41:1864–1886.

[14] Feng R, Jia Y, Wang T, Gan H. Research on the system design and target recognition method of the rebar-tying robot. *Buildings* 2024, 14:838.

[15] Yu Z, Zhang Y, An Q. Measuring the industry golem location and posture repetition exactitude, according to the improved sparrow search arithmetic-improved Gaussian curve matching-improved Hough transform arithmetic. *Measurement* 2025, 242:115976.

[16] Yuan X, Liu G, Sun B, Li Y. A contactless measurement instrument based on fusion of a single-point laser range finder and a VIO system. *J. Phys. Conf. Ser.* 2021, 1820:012112.

[17] Bao J, Li J. Study on short distance laser ranging system. In *The International Conference on Photonics and Optical Engineering*, Xi'an, China, October 13–15, 2014, pp. 874–79.

[18] Ding C, Wang J, Zhong H, Wang Z, Deng L, *et al.* Multimodal dynamic feature fusion framework for human-robot audio-visual interaction in complex construction conditions. *Autom. Constr.* 2026, 187:106951.

[19] Song J, Lu T, Wang H, Zhou J, Zhu Q, *et al.* Automatic rebar spacing measurement using YOLOv11-rebar-light and simplified Hough transform. *Autom. Constr.* 2026, 186:106907.

[20] Tang S, Zhou S, Tong X, Gao J, Wang Y, *et al.* Line segment detectors with deformable attention. *Signal Process. Image Commun.* 2024, 128:117155.

[21] Ji J, Wang M, Yang J, Ren Y, Zhao L. A visual-inertial localization algorithm for greenhouse robots based on point-line features. *Smart Agric. Technol.* 2025, 12:101553.

[22] Bachiller-Burgos P, Manso LJ, Bustos P. A variant of the Hough Transform for the combined detection of corners, segments, and polylines. *EURASIP J. Image Video Process.* 2017, 2017:32.

[23] Li C, Zhang W, Li F, Guo M, Fan S. Vision-based detection, localization, and optimized path planning for rebar intersections in automated construction. *Appl. Sci.* 2025, 15:7186.

[24] Gong L, Zhang F. Investigation on the feasibility of lung lesion detection in CT images for COVID-19 using CG-YOLOv5. *Biomed. Signal Process. Control* 2026, 121:110352.

[25] Yu H, Liu J, Lin M. A comprehensive literature review on YOLO-based small object detection: methods, challenges, and future trends. *Comput. Mater. Contin.* 2026, 87(1):7.

[26] Xu K, Lu X, Shen T, Zhu X, Wang S, *et al.* Rebar binding point location method based on improved YOLOv5 and thinning algorithm. *Measurement* 2025, 242:116029.

[27] Xu S, Guo J, Wang H, Deng L, Shi P, *et al.* Feature-aware autonomous camera pose estimation for rebar tying robots. *Autom. Constr.* 2026, 188:107000.

[28] Li Z, Zheng S, Xiao G, Wu X, Lin B, *et al.* Automatic rebar binding method for PC components based on RGB-D information. *Int. J. Humanoid Robot.* 2026, 23(3):2540007.

[29] Moscatelli M, Tanelli M, Demir AG. Deep learning-powered vision system for seam-tracking and gap width estimation in robotic laser welding. *J. Adv. Joining Process.* 2026, 13:100377.

[30] Li Y, Xiao Z, Li J, Shen T. Integrating vision and laser point cloud data for shield tunnel digital twin modeling. *Autom. Constr.* 2024, 157:105180.

[31] Zhang S, Mohammed A.H. A, Mokhtar N, Nukman Y, Chen H, *et al.* An integration of BCI and vision system for robotics applications: a review. *Chin. J. Mech. Eng.* 2026, 39:100188.

[32] Zheng Y, Liu W, Han L, Li J, Ding H, *et al.* The visual–laser fusion measurement methodology for robotic in situ machining of large components with complex features. *Robot. Comput. Integr. Manuf.* 2026, 100:103257.

[33] Wang C, Dai W, Chen J, Xu X, Ding T. Real-time micrometer-level dynamic pose measurement and uncertainty propagation chain modeling of robot end-effectors based on laser-event hybrid sensing framework. *Measurement* 2026, 290:122775.

[34] Su R, Li J, Liu C, Wang L, Gao L, *et al.* Decoupled active in-row following strategy for collaborative tobacco-harvesting robots based on ultra-wideband localization and visual perception. *Comput. Electron. Agric.* 2026, 253:112152.

[35] Qi J, Xiong Y, Zou Z, Wang S, Wang F, *et al.* Visual–tactile fusion for real-time weed detection in maize: Lightweight precision weed detection network with tactile-triggered keyframes. *Artif. Intell. Agric.* 2026, 16(3): 1835–1851.

[36] Chen H, Wen X, Liu Y, Zhang H. Research on comprehensive vehicle information detection technology based on single-point laser ranging. *Sensors* 2025, 25(5):1303.