scientific reports

Check for updates

OPEN

An intelligent emulsion explosive grasping and filling system based on YOLO-SimAM-GRCNN

Jiangang Yi[1,2,3,6], Peng Liu[1,3,6], Jun Gao[1,3]✉, Rui Yuan[1,2,4,5] & Jiajun Wu[3]

For the blasting scenario, our research develops an emulsion explosive grasping and filling system suitable for tunnel robots. Firstly, we designed a system, YOLO-SimAM-GRCNN, which consists of an inference module and a control module. The inference module primarily consists of a blast hole position detection network based on YOLOv8 and an explosive grasping network based on SimAM-GRCNN. The control module plans and executes the robot's motion control based on the output of the inference module to achieve symmetric grasping and filling operations. Meanwhile, The SimAM-GRCNN grasping network model is utilized to carry out comparative evaluated on the Cornell and Jacquard dataset, achieving a grasping detection accuracy of 98.8% and 95.2%, respectively. In addition, on a self-built emulsion explosive dataset, the grasping detection accuracy reaches 96.4%. The SimAM-GRCNN grasping network model outperforms the original GRCNN by an average of 1.7% in accuracy, achieving a balance between blast holes detection, grasping accuracy and filling speed. Finally, experiments are conducted on the Universal Robots 3 manipulator arm, using distributed deployment and manipulator arm motion control mode to achieve an end-to-end grasping and filling process. On the Jetson Xavier NX development board, the average time consumption is 119.67 s, with average success rates of 87.1% for grasping and 79.2% for filling emulsion explosives.

With the development of the global economy and technological advancements, mine safety has become a focus of attention for countries worldwide. Many countries have started applying intelligent technology in mining production to enhance safety and efficiency. Intelligent robot systems are the core of mining automation. These systems possess functions such as self-perception, decision-making, and control, enabling them to perceive the mining environment, predict and assess risks, and autonomously control the production process. By utilizing intelligent robot systems, manpower can be reduced, production efficiency can be improved, and accident risks can be effectively lowered[1]. "Drilling, blasting, filling, and transportation" are the four key stages in tunnel construction, with the blasting stage lagging in terms of intelligent automation compared to the other three stages. This has become a pressing demand for improving tunnel excavation efficiency. However, the filling of explosives is usually done with manual or makeshift filling equipment. This working environment not only has problems such as humid air and harsh conditions but also imposes heavy labor intensity and extreme danger on workers. Therefore, many large equipment companies are racing to develop emulsion explosive filling equipment.

Our research endeavors to solve an important scientific challenge: how can robots be employed to accomplish the task of intelligent emulsion explosive grasping and blast hole filling in complex tunnel drilling environments? Vision-based object detection and robot grasping techniques have emerged as important research directions in this field. These two technologies have been integrated into end-to-end automation systems for robots and are widely applied in various scenarios, including automobile assembly, fruit picking, and material stacking[2–5]. Vision-based object detection for robots can be achieved using traditional computer vision methods or deep learning like Convolutional Neural Networks (CNN)[6] and You Only Look Once (YOLO)[7]. In this study, we focus on the detection of blast holes positions for robots, and we compare different YOLO series detection algorithms using a self-built blast holes dataset. Based on maintaining accuracy as much as possible, we found that YOLOv8[8] demonstrates significantly improved speed in blast holes detection[9]. Visual grasping with deep learning is a topic that has been diffusely studied in the field of robotics. The process of grasping detection

[1]State Key Laboratory of Precision Blasting, Jianghan University, Wuhan 430056, China. [2]Hubei Key Laboratory of Industrial Fume and Dust Pollution Control, Jianghan University, Wuhan 430056, China. [3]School of Smart Manufacturing, Jianghan University, Wuhan 430056, China. [4]Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan 430081, China. [5]Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan 430081, China. [6]These authors contributed equally to this work: Jiangang Yi and Peng Liu. ✉email: gaojun407104739@163.com

involves utilizing visual sensors to gather precise and swift information about the grasp box, such as the location and orientation of the intended object. Recently, substantial research has been focused on utilizing deep learning techniques that utilize 2D images as input for grasping detection, resulting in substantial theoretical and practical progress[10]. In this study, we present A Simple Attention Module (SimAM), a 3D attention mechanism-based method for detecting grasps in emulsion explosives[11] and a Generative Residual Convolutional Neural Network (GRCNN)[12]. We enhance and optimize the grasping network of the model in 3D space, utilizing the local Mish activation function for propagation. The enhanced SimAM-GRCNN algorithm converts the emulsion explosive grasping detection task into target localization and angle classification, producing outcomes that depict the 2D grasp box. The key contributions of this study can be summarized as follows:

1. A grasp-and-fill system has been developed for the blasting robot to handle emulsion explosives in tunnel scenarios. The system is composed of two modules: an inference module and a control module. The inference module incorporates a YOLOv8-based blast hole position detection network, and a SimAM-GRCNN-based explosive grasping network; the control module plans and executes the robot's motion control based on the detected blast hole positions and emulsion explosive grasping poses to achieve symmetrical grasping and filling operations.
2. An improved grasping model, SimAM-GRCNN, has been developed and integrated with an eye-in-hand calibration method to meet the end-to-end grasping requirements of robots. This novel model, SimAM-GRCNN, achieves significant accuracy improvements over state-of-the-art methods for both single-object and multi-object grasping tasks. Its performance has been rigorously evaluated on well-established public datasets, including Cornell and Jacquard, as well as on a self-built emulsion explosive dataset.
3. The emulsion explosive grasp-and-fill system is deployed on the Jetson Xavier NX development board, leveraging the YOLO-SimAM-GRCNN system. This integration allows for the execution of tasks such as blast holes detection, emulsion explosive grasping, and filling. The average time for a complete grasping and filling process is 119.67 s, with average success rates of 87.1% for grasping emulsion explosives and 79.2% for filling respectively.This work is organized as follows: in the Methodologies section, we present the visual grasp and detection techniques employed by robots. In the System Model section, we establish the model for emulsion explosive filling robots. The Experimental Verification section assesses the effectiveness of YOLOv8 in detecting visual elements through experimentation with the blast holes dataset. We also train and evaluate the SimAM-GRCNN model using three different grasp datasets. In the Grasping and Filling Experiment section, we integrate the system model to create an experimental platform and conduct experiments on explosive grasping and filling using the YOLO-SimAM-GRCNN system. Finally, the Conclusion section summarizes the paper and provides relevant discussions.

Related work

Robot vision detection and grasping are critical research areas in robotics. By combining machine learning methods, robots can achieve accurate perception and identification of target objects, enabling precise target detection and grasping tasks. Meanwhile, research in robot vision detection is also crucial, involving the detection and analysis of target object shape, quality, surface features, etc[13,14]. In the field of robot vision grasping, many researchers are dedicated to developing high-precision target detection and tracking algorithms, enabling robots to accurately locate and track target objects for precise grasping actions[15]. These endeavors require the fusion of knowledge and methodologies from various fields namely computer vision and deep learning. By means of ongoing research and innovation, endeavors are being undertaken to enhance the visual detection and grasping proficiencies of robots to cater to diverse domains, and to facilitate the extensive integration and utilization of robotic technology in pragmatic settings.

Visual detection

Presently, deep learning-based object detection methods can be classified into two categories based on their design. The first type, such as the Region-CNN series (R-CNN), is a two-stage algorithm that utilizes region proposals. The second type of deep learning-based object detection methods is a single-stage algorithm, represented by the Single Shot MultiBox Detector (SSD) and You Only Look Once (YOLO) series, which employs regression[16,17]. In the domain of blast holes detection, several deep learning techniques have been employed. Zhang et al.[18] improved the SqueezeNet and ResNet-51 networks to construct an enhanced Faster R-CNN model for detecting multiple blast holes and challenging blast holes. Through comparative experiments involving multi-scale and multi-level feature fusion of blast holes images and distance-constrained Non-Maximum Suppression (NMS), they achieved fast and efficient recognition and localization of single blast hole. By fusing multi-level features and using distance-constrained NMS for filtering multiple challenging blast holes, they achieved a high accuracy and recall rate in the recognition and localization of multiple blast holes and challenging blast holes. Zhang et al.[19] also used the R-CNN, Faster R-CNN, YOLOv2, and SSD512 network architectures to construct an underground blast holes recognition algorithm. The results showed that YOLOv2 and SSD512 were faster in detection speed compared to Faster R-CNN, but not as accurate. Yue et al.[20] designed the blast holes intelligent detection model MCIW-2 and proposed an anchor box aspect ratio clustering algorithm and a lightweight blast holes intelligent detection model. However, the deployment validation of this model in actual filling tasks has not yet been carried out. YOLO models, ranging from YOLOv5 to YOLOv10, have become increasingly popular in the field of visual detection due to their high accuracy and real-time performance. Researchers have progressively adopted YOLOv8, which builds upon the advancements of previous versions, for applications in robotic vision and grasping[21–24]. This trend is driven by the need for efficient and accurate object detection in dynamic environments, where robots must quickly identify and manipulate objects. YOLOv8, with its robust object detection and segmentation capabilities, offers a powerful tool for enabling robots to perceive their

surroundings and perform complex tasks with precision and speed. Kolin et al.[25]firstly compared YOLOv8-seg with classical methods for object segmentation in robotic grasping. Yan et al.[26]improved YOLOv8s for apple detection and branch/trunk segmentation in modern orchards, enhancing robotic picking. They collected annotated images, augmented the dataset, and integrated SE modules and a dynamic snake convolution module. The improvements led to significant performance gains over YOLOv8s, YOLOv8n, and YOLOv5s. Nevertheless, the algorithm was not tested on an actual robotic picker.

Robot grasping

To achieve stable grasping tasks, researchers first studied the mechanical characteristics and the movement of the end-effector while in contact with objects and conducted grasp analysis as described in previous studies[27,28]. Previous research has used supervised learning methods to handle robot grasping tasks of novel objects by training with synthetic data. However, these methods have limitations in specific environments like offices and kitchens. In order that overcome this challenge, Satish et al.[29] introduced a technique known as Fully Convolutional Grasp Quality Convolutional Neural Network (FC-GQ-CNN). This method utilizes data collection strategies and synthetic training environments to predict stable grasp quality. As a result, it significantly enhances the processing capacity, allowing for a higher number of grasps to be analyzed per second. Furthermore, with the widespread availability of affordable RGB-D cameras. Present research depends on using RGB-D image data to predict grasp poses, recent experiments have shown the effectiveness of deep neural networks, with these methods relying solely on deep learning can efficiently compute stable grasps[30–32]. Mousavian et al. presented a GraspNet with 6 degrees of freedom (6-DoF). This network evaluates the quality of 6D gripper poses by mapping observed target and robot gripper point clouds. Moreover, GraspNet's gradients can guide the robot gripper to prevent collisions and align with the object during manipulation[33]. Murali et al.[34] proposed a technique to plan 6-DoF grasps for objects in cluttered environments using partial point cloud data. Their method enables efficient grasp sequences to be generated for objects that are currently inaccessible . Kumra et al.[35] presented A Real-Time Multi-Grasping Detection Network for Robotic Grasping (GRCNNv2). The network is based on image depth information and extracts multiple potential grasp points in the image using methods such as pyramid dilated convolutions and multi-resolution receptive fields, enabling accurate and efficient grasp point detection in real-time scenarios. Ge et al.[36] arranged a robot grasping method that utilizes a 3D detection network. Using a CNN, they calculated 3D bounding boxes and generated a strategy for optimal grasping poses. However, the network did not incorporate depth information from the camera for fusion. Bin et al.[37] integrated an CBAM attention mechanism into the SqueezeNet architecture, harnessing a sophisticated five-parameter scheme to encode 2D grasp configurations. This innovation facilitated object grasping without necessitating intricate enhancements to the network's design, thereby maintaining architectural simplicity. On another front, Yang et al.[11] introduced the SimAM attention module, distinguished by its streamlined design and remarkable efficacy, further advancing the realm of attention-based methodologies in neural networks.

Different from previous works, our study focuses on visual detection and robotic grasping simultaneously. Table 1 provides a comparative summary of our research findings and the latest trends in robotic grasping and filling.

Problem formulation

In this work, we define the problem of emulsion explosive grasping localization from image to robot grasp transformation in a robotic grasping scenario. In the general object-grasping task of robots, Kumra et al.[35] proposed a grasping model based on eye-to-hand. In this system, an improved version of the eye-in-hand grasping model is proposed, representing the grasping poses in the robot framework as:

$$G_r = (P_r, W_r, \Theta_r, Q_r) \tag{1}$$

The grasping pose of the image can be defined as the center position of the grasping target $P_r = (x_r, y_r, z_r)$, As W_r is the width of the gripper opening necessary for grasping the target, Θ_r is the rotation angle of the tool around the z-axis, and Q_r is the quality score that assesses the current target's grasping quality. We detect grasps from an n-channel image, the image's height and width are represented by *h* and *w*, respectively. The grasping pose of the image can be represented as:

Authors	Robot grasping	Robot filling	Cornell dataset	Jacquard dataset	Explosive dataset
Lenz et al.[2]	✓	×	✓	×	×
Redmon et al.[38]	✓	×	×	×	×
Zhou et al.[39]	✓	×	✓	×	×
Morrison et al.[40]	✓	×	✓	×	×
Yu et al.[41]	✓	×	✓	✓	×
Kumra et al.[35]	✓	×	✓	✓	×
Bin et al.[37]	✓	✓	✓	✓	×
Ours	✓	✓	✓	✓	✓

Table 1. A comparison of related work.

$$G_i = (x_i, y_i, W_i, \Theta_i, Q_i) \tag{2}$$

where (x_i, y_i) is these grasping poses include an image of the grasping target's optimal center coordinates, and W_i is the target width required for grasping, which ranges between $[W_{min}, W_{max}]$ pixels. Basically, Q_i represents the grasping quality of each target center in the image, with values closer to 1 meaning a higher likelihood of success grasping. Θ_i is the rotation angle represents the relative rotation angle concerning the image's x-axis and ranges between $[-\frac{\pi}{2}, \frac{\pi}{2}]$ radians. Therefore, a pose transformation is required to convert the grasping poses of the image to the grasping poses in the robot framework.

Approach

The complex environment of the emulsion explosive grasping and filling robot operation is unique, with variations in the position of the blast holes and the morphology of the channels, and the working environment is in a confined space tunnel. These filling characteristics determine that the structure of the explosive filling robot is targeted[42]. Before the explosive filling, it is necessary to detect the position of the blast holes and the grasping pose of the explosive[43,44], and generate labels corresponding to each blast holes and emulsion explosive based on the results of the front and rear detection, and then plan the motion control route of the robot, and finally execute the mechanical arm to complete the grasping and filling. The proposed YOLO-SimAM-GRCNN system is used for the grasping and filling of emulsion explosives of the robot. The model consists of three main components: the grasping and filling system, the SimAM-GRCNN, and the motion control system. The effectiveness of explosive grasping depends on the accuracy of both the SimAM-GRCNN detection algorithm and the YOLOv8 blast holes detection in the grasping and filling system. The position accuracy of robot motion control affects the success rate of explosive filling.

Grasping and filling system

The grasping and filling system mainly consists of an inference module and a control module, which are used to achieve perception, planning, and decision-making to complete the tasks of blast holes detection and explosive grasping. There are two main parts to the inference module. One component detects blast holes within the camera's field of view, while the other predicts suitable grasping poses for objects within the same field of view. The control module then plans and executes the robot's motion control based on the position of the blast holes and the grasping poses of the emulsion explosive, to achieve symmetric grasping and filling operations.

Inference module

The inference module is comprised of two key components: the blast holes detection network and the emulsion explosive grasping network, as illustrated in Fig. 1. The RGB-D camera captures the scene's image, which is then preprocessed to align with the network's input requirements. The blast holes detection network is the first

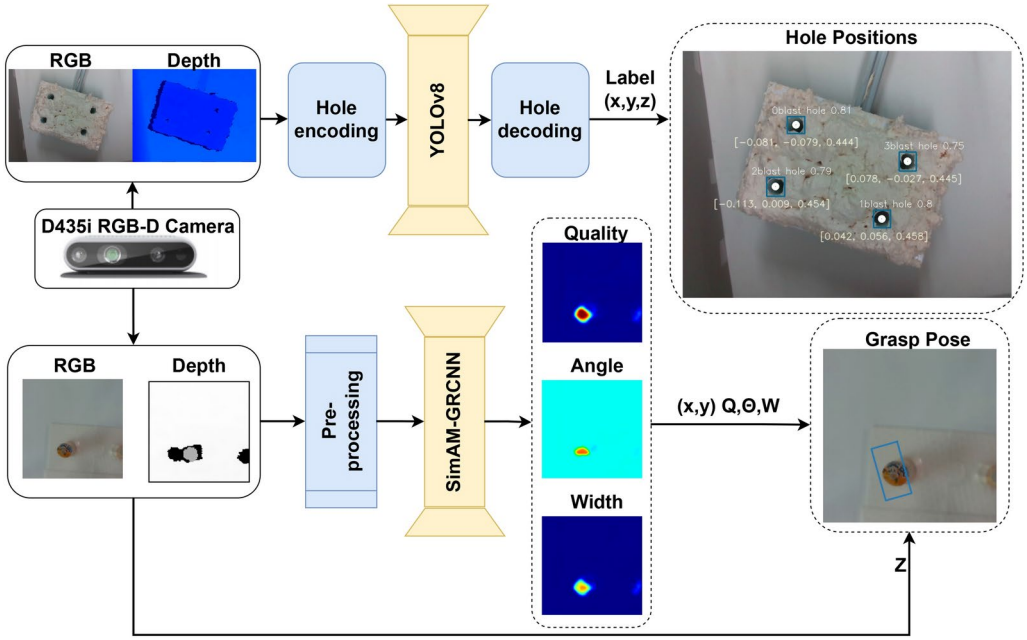

Fig. 1. System inference module. The inference part of the system consists of two components: the blast hole detection model and the grasping detection model. Both models rely on RGB-D images captured by the D435i camera after the UR3 robotic arm moves to specified positions. The output of the upper blast hole detection model is a series of blast hole filling locations, while the output of the lower grasping detection model is a series of explosive grasping positions.

component of the inference module. By normalizing and encoding the feature images, it preprocesses RGB-D four-channel images. A YOLOv8 network inference is then performed on the RGB image to obtain the blast hole centers. An estimate of the hole depth is obtained by averaging the depth information of the four corners of the YOLOv8 detection bounding boxes, with the holes surface depth obtained from the average depth value and the compensation depth representing the simulated holes depth[45,46]. The depth of the blast hole center position is calculated based on the compensation depth. According to the YOLOv8 blast holes detection network, the blast hole numbers and their corresponding three-dimensional positions are listed in Fig. S1 in the supplementary section. Finally, the compensatory depth calculated above is converted into the final fill position for the target hole. The other component is the emulsion explosive grasping module, It consists of three parts: (1) Input data is preprocessed, including cropping, resizing, and normalization. The processed image is fed into SimAM-GRCNN. SimAM-GRCNN can use inputs of any channel number, not limited to specific types of input modalities, making it versatile for any type of input modality. (2) The SimAM-GRCNN network extracts features from the preprocessed image and generates three feature maps as outputs: one for the grasping angle, another for the grasping width, and a third for the grasping quality score. (3) The grasping pose is inferred based on these three output images. By combining the detection results from both components, we can generate one-to-one labels that match the blast holes filling positions with the emulsion explosive grasping positions. This completes the end-to-end position matching of the robot. Additionally, we need to implement robot motion control to plan and execute the motion control routes of the mechanical arm for grasping and filling, ensuring the successful completion of the matched actions.

Control module

There is mainly a task controller in the control module, which uses the inference module to generate the starting point grasping pose and the endpoint filling position of Robot's end-effector determines the system's motion tasks. The control tasks of the system are implemented in an orderly manner through the Python interface of the trajectory controller and planner, as shown in Fig. 2.

The task controller carries out operations such as grasping, filling, and calibrating the control module. It requests grasping poses from the inference module, which then delivers them in descending order of quality. By matching them one by one with the previously obtained blast holes information, multiple sets of robot starting positions can be obtained. Furthermore, the Python interface is used to perform inverse kinematics trajectory planning for the robot's end-effector actions. 7-DOF robot enables the trajectory controller based on the starting point and executes the planned trajectory to perform the corresponding grasping and filling tasks. Due to the adoption of a modular design approach and integration with Python, this system is suitable for most robots available on the market.

SimAM-GRCNN

SimAM-GRCNN is a model that generates pixel-level grasp results from an input image with 4-channels as shown in Fig. 3. Three convolutional layers and a channel attention SimAM module are applied to the 4-channel image in the first step. Then, it goes through five residual layers and another channel attention SimAM module. Next, three convolutional layers are applied. There are four tensors included in this set: grasp quality score, trigonometric function of angle $\sin 2\Theta \& \cos 2\Theta$ with plane, and end-effector width. The grasp angle is obtained by $Angle = \arctan \frac{\sin 2\Theta}{\cos 2\Theta}/2$ calculating the angle relative to the X-axis[47], and finally, the optimal grasp box is generated.

Network structure

Firstly, the SimAM-GRCNN grasping network utilizes three convolutional layers to extract initial features from the input RGB-D image. The first improvement made in the SimAM-GRCNN grasping network is the incorporation of the Mish activation function to replace the traditional ReLU activation function in the first Conv2D + BatchNorm + Activation (CBA1) module. It improves training stability due to the slight negative values allowed by the Mish function, which facilitate better gradient flow compared to the hard zero boundary of ReLU. Furthermore, a plug-in SimAM11 module incorporating a 3D attention mechanism is integrated before and after the Residual Block module. Compared to existing channel and spatial attention modules, SimAM can infer 3-D attention weights for feature maps within a layer without adding parameters to the original network. Another advantage of this module is that most operators are chosen as solutions to a defined energy function, which avoids too much structural adjustment work. The features refined by the SimAM module are better

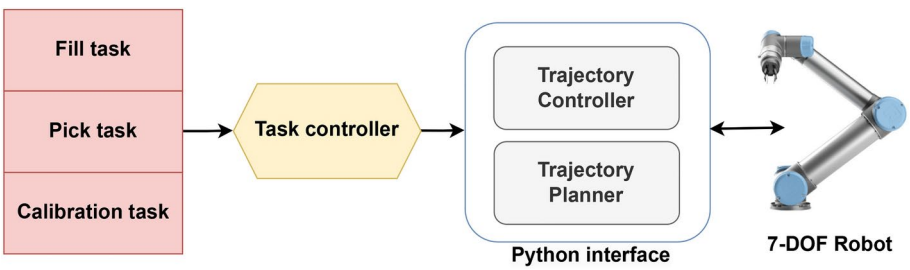

Fig. 2. System control module.

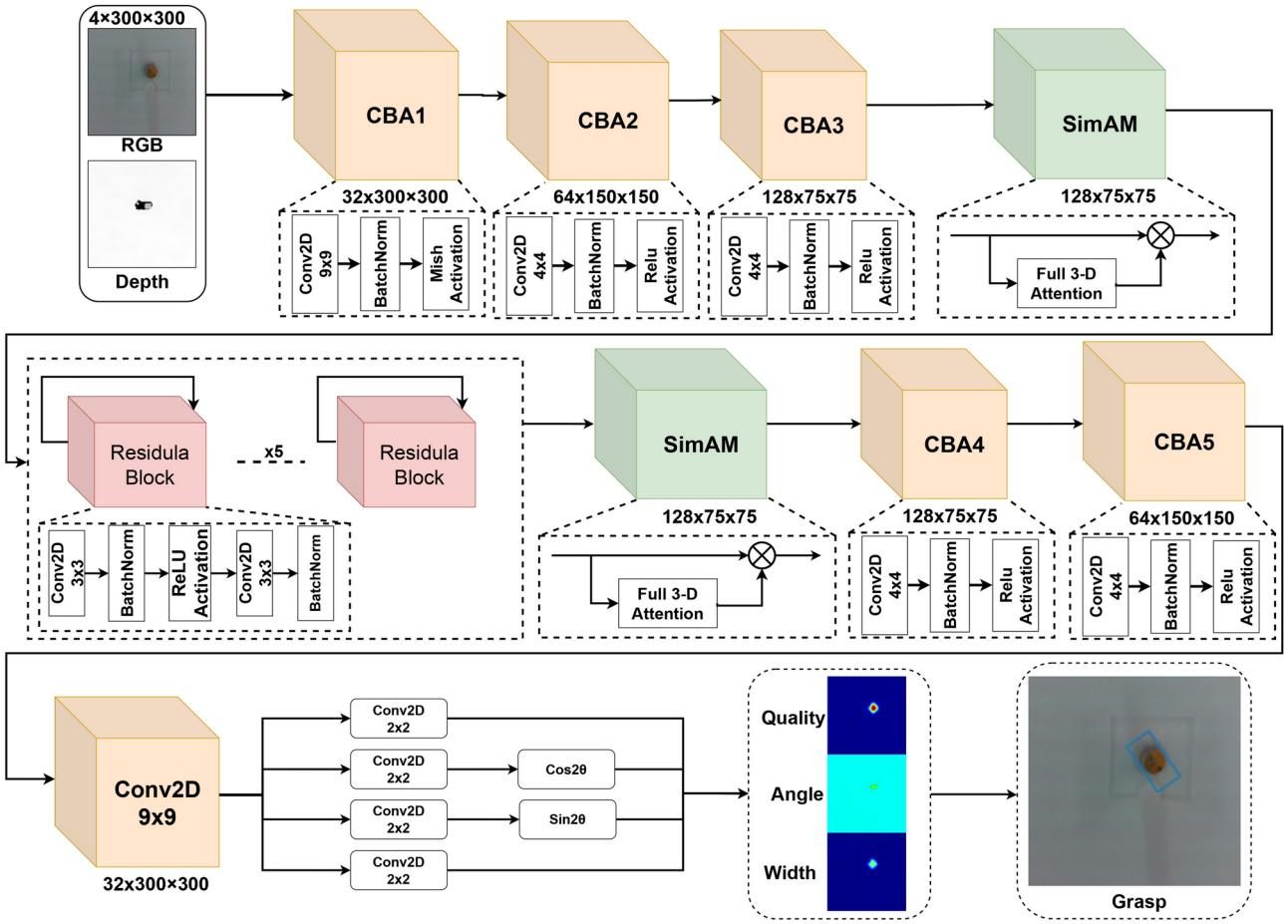

Fig. 3. SimAM-GRCNN network structure. The SimAM Generative Residual Convolutional Neural Network (SimAM-GRCNN) comprises three main components: the input stage, the backbone network, and the output stage. First, the input data undergo preprocessing, which includes cropping, resizing, and normalizing the images to 300×300 pixels with 4 channels (RGB-D). Second, the backbone network extracts grasp features from the preprocessed images. Finally, the output stage infers the grasp pose data for the target object in the image, including quality, angle, and width.

able to concentrate on the target object for grasping. When the number of convolutional layers in a network structure exceeds a certain threshold, accuracy no longer improves and issues such as gradient vanishing and dimensionality errors may arise. Thus, the grasp network utilizing 5 residual layers with skip connections enhances the learning of RGB-D features more efficiently. After processing through these convolutional and residual layers, the image size is reduced to 75×75. To facilitate interpretation and to retain spatial features of the image after convolutional operations, the grasp network utilizes transposed convolutional operations to upsample the image, aligning it with the original input size to obtain an output image of the same dimensions. The SimAM-GRCNN grasp network has a total of 1.9 MB of parameters, which is equivalent to the parameter size of the original GRCNN grasp network. The SimAM-GRCNN grasp network proposed in this study has fewer parameters and a lower computational cost compared to other grasp networks. This model is suitable for robot closed-loop grasp motion control at up to 50 Hz.

Training methodology

In the training process of the grasping network, we employ the conventional backpropagation algorithm as our training strategy. The network architecture integrates both Adam and SGD optimizers to explore their distinct effects on model training. The learning rate is set to 10^{-3}, and dropout regularization is systematically applied to enhance the model's generalization capabilities. The grasping network training utilizes a small batch size of 8, with each epoch consisting of 1,000 such batch iterations. All experiments are consistently executed under a fixed random seed of 123 to ensure the reproducibility of the results. Comparative analyses reveal that Adam optimizer demonstrates superior robustness compared to SGD during training, rapidly and efficiently converging to favorable model parameter configurations, thereby hastening the progression toward the optimal solution.

Loss function

Loss functions of L1, L2, and Smooth L1 were analyzed. Following the training of the SimAM-GRCNN grasping network, it was observed that the former two loss functions encountered issues with gradient explosion in later stages of training, whereas the Smooth L1 Loss did not exhibit such explosions, demonstrating enhanced robustness in handling outliers. We found that Smooth L1 Loss has less fluctuations when handling outliers in the SimAM-GRCNN grasping network. $L(G_i, \hat{G}_i)$ is the loss function, which quantifies the difference between the predicted grasp results G_i and the actual grasp results $\hat{G}_i$:

$$L(G_i, \hat{G}_i) = \frac{1}{N} \sum_{i=0}^{N-1} SmoothL1(y_i - \hat{y}_i) \tag{3}$$

where *i* indicates the index of the *i*-th grasp bounding box, and *N* is the total number of grasp predictions being evaluated. y_i represents the grasp results generated by the network, while $\hat{y}_i$ represents the actual grasp results. where SmoothL1 is given by:

$$SmoothL1(x) = \begin{cases} \frac{1}{2}(x)^2 & \text{if } |x| < 1 \\ |x| - \frac{1}{2} & \text{otherwise} \end{cases} \tag{4}$$

The overall loss function L represented in Eq. (3) encompasses the losses of the image outputs. A model's combined loss of quality $L_{quality}$, desired width L_{width},sine angle L_{sin} and cosine angle L_{cos} is calculated as follows:

$$L = L_{quality} + L_{width} + L_{sin} + L_{cos} \tag{5}$$

Pose transformation

To achieve coordinated control between the camera and the robot for grasp poses, hand-eye calibration is necessary. Hand-eye calibration is the procedure used to establish the correlation between the camera coordinate system and either the end-effector coordinate system or the robot base coordinate system[48,49]. When modeling the working environment of a mobile robot, Xiao et al.[50] recommends the "eye-in-hand" method, where both the camera and the tool are installed on the robot's end-effector. As a consequence of this situation, the manipulator and target object are getting closer, which requires higher stability of the visual camera. As a result, the absolute error of the measured target position parameters decreases. "Eye-to-hand" refers to the visual camera installed in a fixed position outside the robot arm, which has a relatively stationary perspective. The disadvantage is that the details in front of the manipulator are difficult to capture when the manipulator is in motion. Compared with the above two calibration methods, this system needs to transform the camera's position to perform holes detection and emulsion explosive grasping pose recognition separately. Therefore, the "eye-in-hand" hand-eye calibration method is adopted, and one camera can meet the above requirements, reducing the number of cameras and system operation costs. Eye-in-hand camera calibration schematic in Fig. S2 in the supplementary section. Accordingly, the grasping pose from the image to the robot can be defined as follows:

$$G_r = T_{Tool}^{Base} * T_{Camera}^{Tool} * T_{Image}^{Camera} * G_i \tag{6}$$

One of the transformations T_{Image}^{Camera} involves converting the image space to 3D space of the camera based on its intrinsic parameters. Then, the transformations T_{Camera}^{Tool} involves converting the camera space is transformed into the tool space using the camera pose calibration values. Finally, the transformations T_{Tool}^{Base} involves converting the tool space pose is transformed into the robot space. By using the eye-in-hand calibration method, the grasp pose $(x_i, y_i, W_i, \Theta_i, Q_i)$ in the image is transformed into the grasp pose $(x_r, y_r, W_r, \Theta_r, Q_r)$ in the robot's space. Then, aligning the depth and color images allows us to obtain the axial depth z_r information of the target object for grasping, thus obtaining the complete grasp pose $(P_r, W_r, \Theta_r, Q_r)$ for the robot. This method can be used for multi-object grasping in images. As a result, we can represent the collection of all grasps as follows:

$$G = (W, \Theta, Q) \in R^{p \times h \times w} \tag{7}$$

An image's grasping width, angle, and quality score are calculated per pixel as $W/\Theta/Q$. The complete set of grasps that map the 3D environment is represented by *R*, *p* represents the three-dimensional position of the center point of the grasping box, The grasp is formed by the combined height *h* and width *w* of the grasp box.

Motion control system

Building upon the key technologies outlined in the Approach section, the system leverages the YOLOv8 model for blast holes detection and employs the SimAM-GRCNN model to detect the grasp points of emulsion explosives. These two detection models jointly facilitate the generation of target positional information for the robot end-effector, thereby enabling the execution of both grasping and filling operations by the UR3 robot. The experimental system is designed to perform end-to-end operations, commencing with the explosives on the charging platform and culminating at the blast holes located on the tunnel face. All the points of robot movement are autonomous and controllable, which ensures the safety and stability of the system's motion control. Combining the end-to-end position information set transmitted by the two models mentioned above, the next step is to interact with the motion control system for execution. This motion control system consists

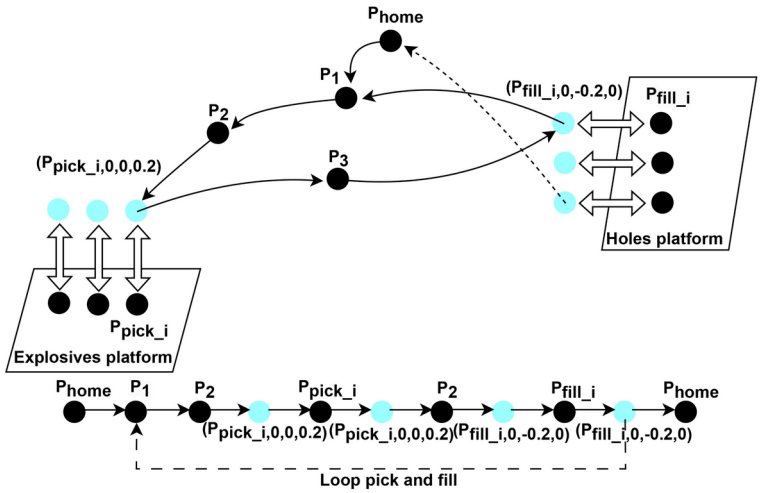

Fig. 4. UR3 robot motion process.

Dataset	Data type	Size	Objects	Images	Grasps	Boxes
Blast holes	RGB	640×480	3	1215	–	1.2k
Cornell	RGBD	640×480	240	1035	8k	–
Jacquard	RGBD	1024×1024	11619	54485	1.1M	–
Explosives	RGBD	640×480	5	3589	32k	–

Table 2. Summary of four datasets. Boxes represent bounding box.

of systems such as visual detection, grasping, and filling, and the entire simplified motion process is shown in Fig. 4.

Visual detection system
First, the robot is powered on and initialized at the P_{home} point. And then it's started at position P_1 in front of the blast holes[51]. Afterward, the robot tool proceeds to the blast holes and captures RGB-D image data. The YOLOv8 visual detection network for blast holes reveals the quantity and positioning of blast holes on a blast hole surface. Next, the robot tool moves to position P_2 above the emulsion explosive to be grasped. The SimAM-GRCNN grasping network is used to recognize the optimal grasping pose for the explosive. This completes the blast holes detection and emulsion explosive grasping detection process.

Grasping system
End-effector of robot moves above grasping point $(P_{pick\_i}, 0, 0, 0.2)$, opens the gripper, and then moves in a straight line to the optimal grasping position P_{pick_i}. As soon as the gripper closes, the robot returns to the position 20 cm above the grasping point$(P_{pick_i}, 0, 0, 0.2)$ to capture the emulsion explosive head. This strategy ensures a smooth linear extraction of the emulsion explosive.

Filling system
After passing through waypoint P_3, which effectively avoids obstacles, the robot reaches a position 20cm in front of the blast holes $(P_{fill\_i}, 0, -0.2, 0)$. At this point, the filling process can begin. Based on the matched emulsion explosive and blast holes positions, the robot moves slowly in a straight line to the corresponding depth position of the blast holes P_{fill_i}. This process represents one complete cycle from blast holes detection to grasping and filling. To fill other blast holes, the robot returns to point P_1 and repeats the process of grasping and filling emulsion explosives. Finally, when all tasks are completed, the robot returns to the P_{home} position.

Experimental verification

Firstly, we need to collect four types of datasets for this experiment. Next, we will set up the experimental environment and connect the relevant hardware to ensure their proper interaction. Finally, we will conduct experimental evaluations of the YOLOv8 blast holes detection model and the SimAM-GRCNN grasping model.

Datasets

We provide an overview of four datasets used in this study Table 2. The YOLOv8 object detection model is evaluated using the first blast holes dataset, whereas our grasping model is trained and evaluated using the last three datasets. The first blast holes dataset is used for training the detection of blast holes center positions, which

differs from the grasping dataset used to train the generation of grasping positions. The second dataset is the Cornell Grasping dataset, which is commonly used as a benchmark for grasping results. Another dataset is the Jacquard Grasping dataset, which is more than 50 times bigger than the Cornell dataset[52]. Thirdly, we present a self-built emulsion explosives grasping dataset that is formatted based on Cornell Grasping dataset.

Blast holes dataset

The data utilized in this research includes images of blast holes obtained from the front of a simulated tunnel project. During the process of collecting image data, the working face was photographed from different angles and under different lighting conditions. The system used three different types of rock masses to extract the blast holes, as shown in Fig. 5.

The ABC rows represent three different blast hole surface features. Individual blast holes images have been enhanced using enhancement techniques such as cropping, rotation, and brightness adjustment. By employing these techniques, we obtained more localized blast hole images improves the robustness of the blast holes detection model as well as enriches the diversity of blast hole dataset.

Emulsion explosives dataset

The emulsion explosives dataset belongs to the grasping dataset. Taking into account the input parameter characteristics of the SimAM-GRCNN grasping model, three types of target grasping information were extracted during the training of the Jacquard grasping dataset: RGB.png, perfect_depth.tiff, and grasps.txt. RGB.png is a PNG image rendered using Blender. perfect_depth.tiff is a float32 tiff depth image obtained by overlaying the PNG image. Grasps.txt is a text file containing target grasping information, with each line containing five parameters that match the parameters in Formula 2. The above parameters can represent a capture rectangle. Once the robot gripper is selected, the grasp rectangle's width is determined. The dataset samples are shown in Fig. 6. To increase the diversity of the dataset, data augmentation was performed by varying the placement angle, and radius size, adjusting brightness, and adding noise, thereby enriching the breadth of information in the emulsion explosive grasping dataset.

Experimental setup

The entire system consists of four components: UR3 robotic arm, robotic arm controller, computer, and binocular vision sensor. The hardware connections of this system are shown in Fig. 7. The binocular vision sensor is used to capture scene image information and transmit it to the computer. In the binocular vision sensor's coordinate system, pose information can be obtained by processing images using image processing algorithms. Through coordinate transformation, the target's pose in the robotic arm's base coordinate system can be determined. Controllers for robotic arms utilize this data to guide the arm in completing the grasping task. During the experimental setup, the scene image was captured using a RealSense D435i binocular depth camera. This camera can generate high-precision RGB-D images in real time and provide low-latency 3D perception capabilities. It also has a wide field of view, allowing for coverage of larger scenes, and supports long-range depth perception. A 7-DOF UR3 with a Robotiq 2F-85 gripper is used. Two fingers are used to grasp the emulsion explosive using a parallel gripper mounted on the robot's tool end. The UR3 controller communicates with controllers A and B via TCP/IP to the host machine. For more details, please refer to the Supplementary Experimental Setup. The controller A is equipped with an Intel Core i5-13400 processor and an NVIDIA GeForce GTX 3060

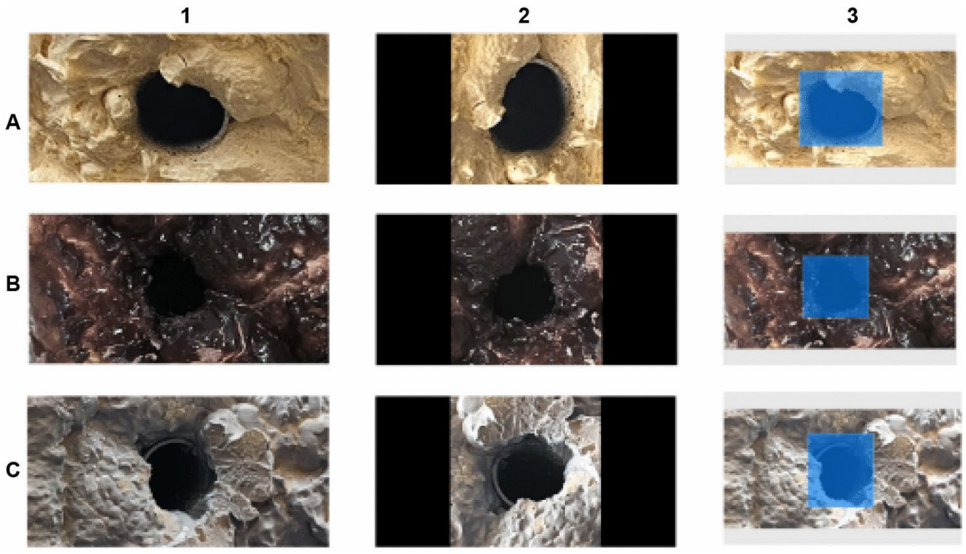

Fig. 5. Visualization of the blast holes part dataset. Row A represents the blast hole surface model in yellow, Row B is the blast hole surface model in reddish-brown, and Row C is the blast hole surface model in silver-gray. Column 1: Original image. Column 2: Image rotated 90 degrees counterclockwise. Column 3: Annotated blast hole image with exposure adjustment.

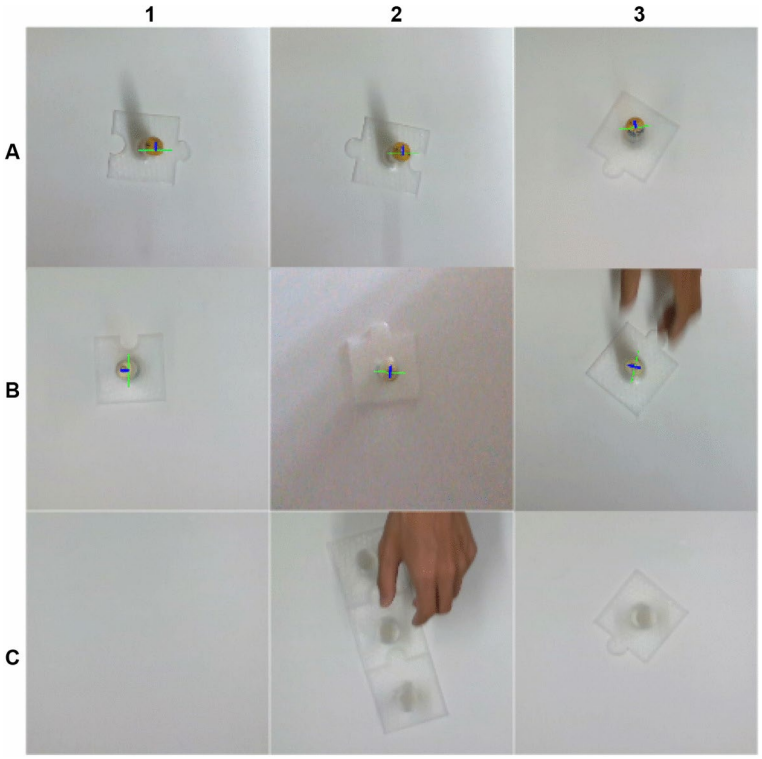

Fig. 6. Visualization of emulsion explosive labeling. Row A consists of the emulsion explosive model in yellow, Row B features the emulsion explosive model in off-white, and Row C is a model with a blank background. Column 123 shows images of emulsion explosives in different positions.

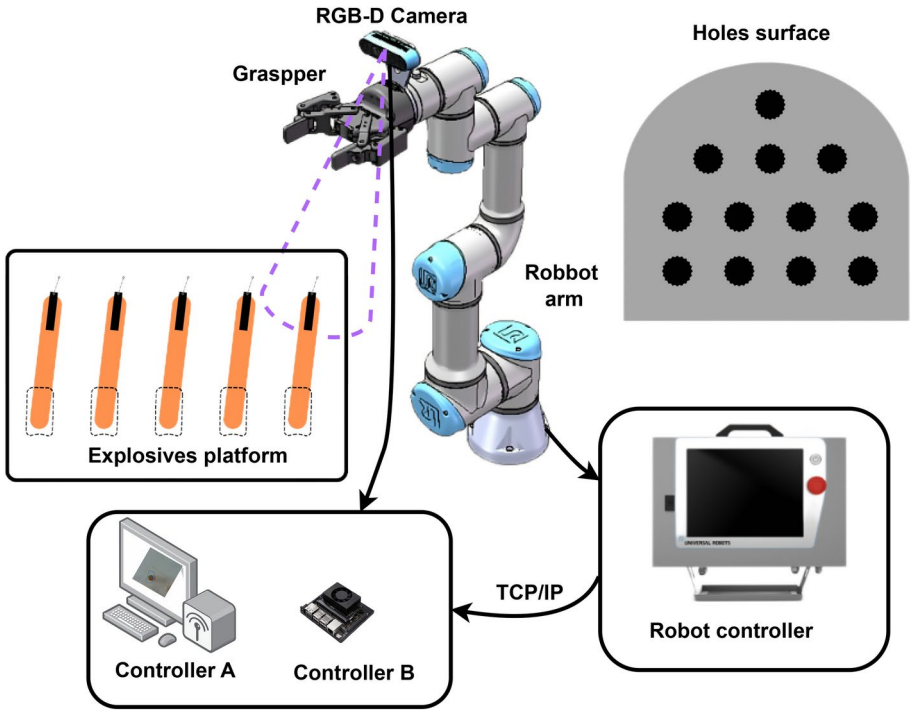

Fig. 7. System hardware connection.

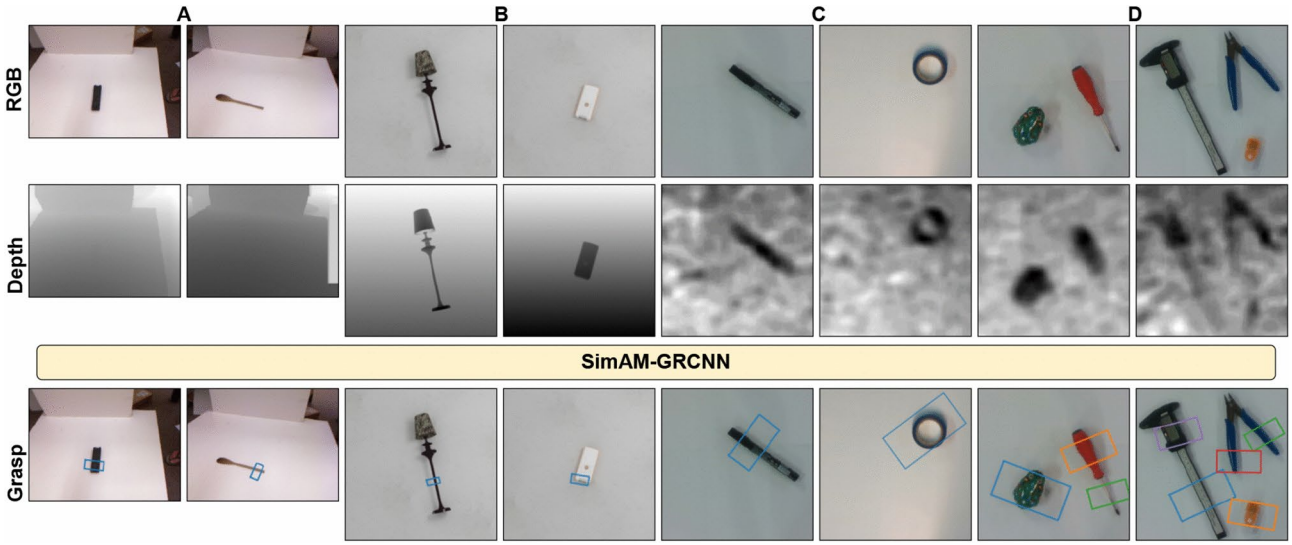

Fig. 8. The performance of the SimAM-GRCNN model on Cornell, Jacquard test sets and everyday life capture images.

Authors	Algorithm	Params (M)	Accuracy (%)	Speed (fps)
Zhang et al.[18]	VGG -16	102.8	88.99	4.17
Zhang et al.[19]	Faster R- CNN	54.3	97.30	5.34
Yue et al.[20]	MCIW-2	**2.8**	96.18	58.72
Ours	YOLOv8	3.2	**97.42**	**76.92**

Table 3. Comparison of blast holes detection models. Bold marking indicates the optimal value under the evaluation criterion

graphics card with CUDA 10 and Python 3.8. Controller B is the Jetson Xavier NX development board, which integrates the Jetpack 5.0.2 system and provides the pytorch environment for fast deployment and experimental verification of this system.

Performance evaluation of models on different datasets

The YOLOv8 model is first evaluated on blast holes dataset, thereafter SimAM-GRCNN is evaluated on Cornell, Jacquard and everyday life tools dataset. As shown in Fig. 8, to convert the grasp representation from image-based to rectangular, it is essential to match the values assigned to each pixel in the output image to its corresponding rectangular coordinates. A. Cornell testing dataset images; B. Jacquard testing dataset images; C. Single-object images in daily life; D. Multi-object images in daily life. According to this paper's rectangle metric, a grasp is valid if it meets the following two criteria: An Intersection over Union (IoU) intersection value greater than 25% and a grasp direction deviation less than 30°. For objects not involved in the training process in the Cornell and Jacquard datasets, our grasping network can reliably generate grasp poses for different types of objects. Furthermore, we demonstrate that the SimAM-GRCNN model can generate multiple grasp poses for multiple objects in complex environments, not only for isolated objects. Finally, we separately evaluate the self-built emulsion explosives dataset.

Evaluation on blast holes dataset

Training and test sets were randomly divided 8:2 during the experimentation with the YOLOv8 blast holes detection model. Blast holes detection is considered valid when the IoU intersection value exceeds 45% and the confidence is 0.6. We evaluated the YOLOv8 detection model using the blast holes dataset from Section 4.1, with the detection results shown in Table 3.

Model parameters are defined by the corresponding row of parameter sizes, and average detection accuracy is derived from the blast holedataset test set. Detection speed is evaluated using the frame rate (fps). By comparing the average detection accuracy and detection speed, we found that YOLOv8 achieves an average precision of 97.42% in blast holes detection, with a detection speed of 76.92 fps. YOLOv8 strikes a balance between detection speed and accuracy, as demonstrated by experimental results. Therefore, we choose to adopt YOLOv8 as the blast holes detection model and use it as the final filling position reference for the robot end, for subsequent grasp model evaluation.

In this study, during the evaluation on blast holes dataset, the center of the tunnel face is considered as the origin, and the angle between the camera and the plane perpendicular to the tunnel face is adjusted to $0°$, $15°$, $30°$, and $45°$ for comparative experiments shown in Fig. 9.

Utilizing the YOLOv8 blast hole detection model, the system captures images from various perspectives to conduct experiments with three types of blast hole surface. A recognition threshold of 0.7 is set to accommodate the presence of inclination angles. The blast hole models inclined up to $30°$ are generally identifiable, with confidence scores surpassing 0.8. When the inclination angle is approximately $45°$, the detection model successfully captures blast holes with excellent formation quality. However, for the blast hole surface in reddish-brown with challenging recognition, the lower quality in detecting these blast holes leads to recognition difficulty. It is evident that under conditions of significant angle inclination, most of the unidentified blast holes are largely obstructed by surrounding protrusions. The comparative results from these experiments illustrate the robustness of the blast holes detection model in handling different inclination angles and colors.

Evaluation on Cornell dataset

The Cornell grasping dataset is a RGB-D based dataset for grasp point detection. The dataset has been randomly divided into training and validating sets in a ratio of 9:1. The dataset consists of RGB images, depth information, positive sample labels, and negative sample labels. The image size is 640×480. The depth information is recorded in a txt format and needs to be converted to aligned tiff format with RGB before being input into the SimAM-GRCNN grasping network. With RGB-D input, grasp rectangles and recognition time are generated, and accuracy is evaluated using a validation set. The comparative evaluation results with other previous models are shown in Table 4. On the Cornell dataset, SimAM-GRCNN grasping model obtained state-of-the-art accuracy of 98.8% on image split and 97.7% on object split with a generation speed of 19 ms per image. Therefore, the SimAM-GRCNN grasping network is suitable for grasping tasks of common objects in daily life.

Evaluation on Jacquard dataset

The Jacquard dataset consists of RGB images, binary segmentation masks of object scenes, two depth images, and annotated grasp boxes. The dataset has been randomly divided into training and validating sets in a ratio of 9:1. For the research work on Jacquard grasping, previous experimental data was collected as shown in Table 4. Due to the large size of the original Jacquard images and the requirement for complete input feature coordination in the SimAM-GRCNN grasping model, we used RGB-D and annotated grasp box data as inputs. According to Table 5, the grasp accuracy of the SimAM-GRCNN grasping model reached 95.2%, outperforming other models.

Evaluation on emulsion explosives dataset

This study randomly divides the self-built grasp dataset into two sets, a training set and a validating set, using a 9:1 ratio. To validate the functionality of attention mechanisms in accelerating model convergence, we plot the

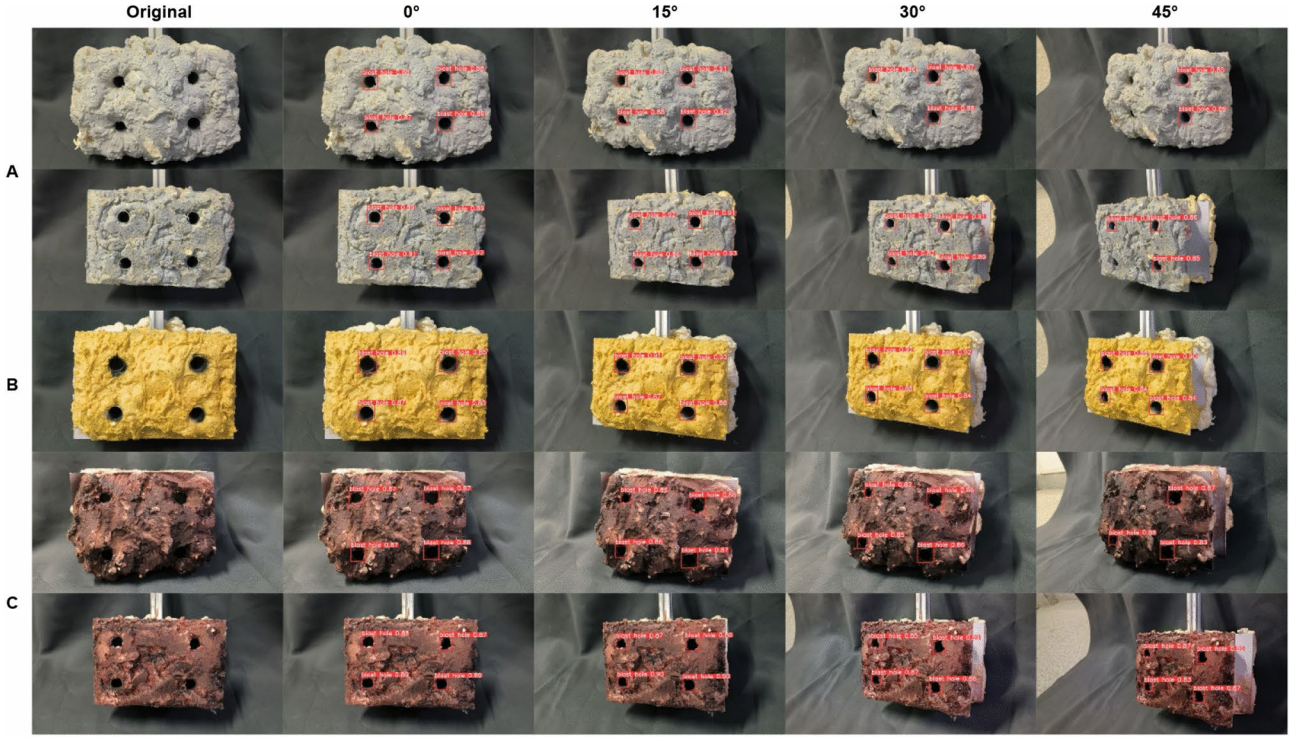

Fig. 9. The test results of YOLOv8 on the face of the biological blast hole at different inclination angles and colors. Row A represents the blast hole surface model in silver-gray, Row B is the blast hole surface model in yellow, and Row C is the blast hole surface model in reddish-brown.

Authors	Algorithm	Image split accuracy (%)	Object split accuracy (%)	Speed (ms)
Jiang et al.[53]	Fast Search	60.5	58.3	5000
Redmon et al.[38]	AlexNet	88.0	87.1	75
Kumra et al.[54]	ResNet-50x2	89.2	88.9	103
Zhou et al.[39]	ResNet-101	97.7	96.6	117
Asif et al.[55]	GraspNet	90.2	90.6	24
Morrison et al.[40]	GG-CNN2	84.0	82.0	20
Yu et al.[41]	SE-ResUNet	98.2	97.1	25
Kumra et al.[35]	GRCNN v2	**98.8**	**97.7**	20
Ours	SimAM-GRCNN	**98.8**	**97.7**	**19**

Table 4. Cornell dataset detection model evaluation. Bold marking indicates the optimal value under the evaluation criterion

Authors	Algorithm	Accuracy (%)
Depierre et al.[52]	AlexNet	74.2
Morrison et al.[40]	GG-CNN2	84.0
Zhou et al.[39]	ResNet-101	92.8
Kumra et al.[35]	GRCNN v2	95.1
Ours	SimAM-GRCNN	**95.2**

Table 5. Jacquard dataset detection model evaluation. Bold marking indicates the optimal value under the evaluation criterion

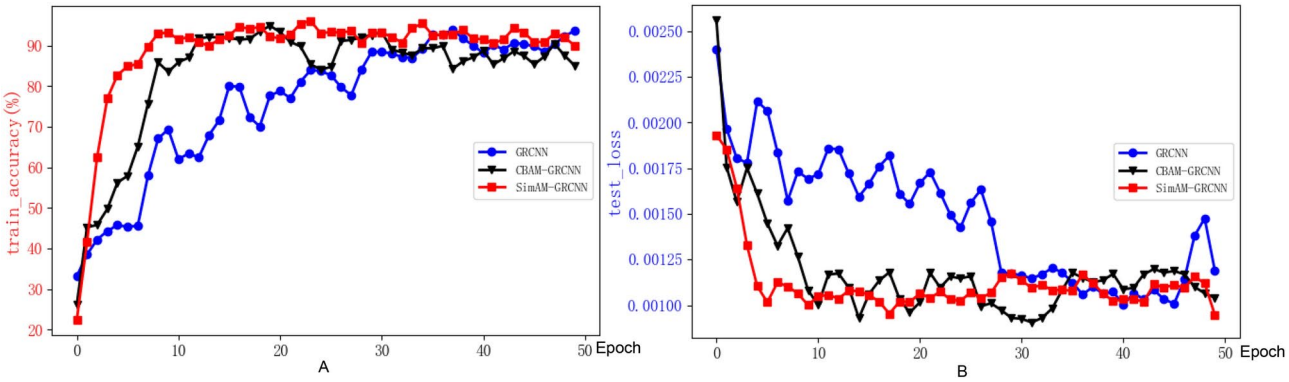

Fig. 10. Three grasping models in emulsion explosive training set accuracy and loss iteration results.A. The variation in accuracy over 50 epochs for three grasping models on the training set; B. The variation of loss rates over 50 epochs for three grasping models on the test set.

changes in inference accuracy over time during the training process. On the emulsion explosive dataset, Fig. 10 shows the grasping detection accuracy and loss for the three grasping models in the supplementary experimental evaluation section(Table S1).

GRCNN is a general grasping model proposed by Kumra, and the other two models incorporate CBAM and SimAM attention mechanism modules into this grasping model, respectively. During the first 7 epochs of training, the two models with attention mechanism modules show a noticeable improvement in training speed, with CBAM being more pronounced. However, at approximately 22 epochs, the SimAM-GRCNN model has already approached the highest accuracy. Introducing the attention mechanism around the 25th epoch accelerates the model's convergence, and a negative correlation is observed between the loss curve on the test set and the accuracy curve on the training set. Therefore, SimAM-GRCNN demonstrates improved accuracy and convergence speed under the same training conditions, indicating that the SimAM attention mechanism improves performance.

In this study, the performance of the trained grasping models is evaluated using common evaluation metrics such as accuracy, precision, recall, and F1 score. The calculation equations for these metrics are as follows:

Accuracy: Measures the proportion of all correctly classified instances (both grasping poses and backgrounds).

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{8}$$

Precision: Measures the proportion of correctly detected grasping poses among all detected grasping poses.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{9}$$

Recall: Measures the proportion of correctly detected grasping poses among all actual grasping poses.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{10}$$

F1 Score: Provides a balanced measure of precision and recall.

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{11}$$

where: TP (True Positives) means the number of correctly detected emulsion explosive grasping poses. FP (False Positives) represents the number of misdetected background items as emulsion explosive grasping poses. FN (False Negatives) means the number of undetected emulsion explosive grasping poses. TN (True Negatives) represents the number of correctly detected background items.

Table 6 summarizes the evaluation results for all three grasp models mentioned above. The SimAM-GRCNN model outperforms the GRCNN and CBAM-GRCNN models in terms of F1 Score. The F1 Score for the SimAM-GRCNN is the highest at 97.8%, which is 0.5% higher than the CBAM-GRCNN and 1.0% higher than the GRCNN. This indicates that the SimAM-GRCNN model achieves a better balance between precision and recall, making it more effective at detecting emulsion explosive grasping poses. Notably, the SimAM-GRCNN also excels in accuracy (96.4%), precision (99.0%), and recall (96.7%) while maintaining a competitive model size (1,901 kilobytes) and processing speed (19 ms). These results suggest that the SimAM-GRCNN model is a superior choice for grasping detection tasks in mining operations, offering a combination of high performance and efficiency. Therefore, the use of the SimAM-GRCNN grasp model enables effective emulsion explosive grasping detection, providing the starting and ending positions for robot end-effector grasping. This, combined with the blast hole dataset section, forms the basis for end-to-end grasping and filling experiments on the robot.

Using our self-built emulsion explosive dataset, we evaluate the SimAM-GRCNN grasping network. As shown in Fig. 11, the study compares the generation of grasp boxes and detection inference based on three grasping networks. In the context of emulsion explosive grasping experiments, "1" following the letters ABC indicates a single emulsion explosive grasping experiment, while "2" signifies multiple emulsion explosive grasping experiments. The experimental evaluation assessed the GRCNN, CBAM-GRCNN, and SimAM-GRCNN models for detecting single and multiple emulsion explosives. The latter two models, which incorporate attention mechanisms, represent enhancements over the first. Models with attention mechanisms, specifically Groups B and C, demonstrate greater sensitivity to the quality and width of the emulsion explosives. The average accuracy of the SimAM-GRCNN grasping network is 1.7% higher than that of the original GRCNN model, and 0.8% higher than that of the CBAM-GRCNN model, while the model's parameter and detection speed are on par with the original GRCNN. In conclusion, the SimAM-GRCNN grasping network can generate stable and accurate grasping postures for both single and multiple emulsion explosives.

Grasping and filling experiment

Based on the key technologies and model deployments mentioned earlier, we successfully established a grasping and filling experiment system. This system utilizes the YOLOv8 model to detect blast holes and the SimAM-GRCNN model to detect the grasp of emulsion explosives. With these models, we can generate position information for the robot gripper to enable grasping and filling operations on the UR3 robot. This experimental system allows for end-to-end operations, covering the entire process from grasping to filling.

Experimental procedure

A dual approach is employed, which includes Grasping detection-based explosive grasping strategy and Object detection-based blast holes filling strategy, both of which are executed in a sequential manner. Subsequently, the primary parameters that affect both explosive grasping and blast hole filling are investigated individually. After

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 score (%)	Params (Kb)	Speed (ms)
GRCNN	94.7	98.3	95.3	96.8	**1901**	20
CBAM-GRCNN	95.6	98.6	96.0	97.3	1905	21
SimAM-GRCNN	**96.4**	**99.0**	**96.7**	**97.8**	**1901**	**19**

Table 6. Comparison of three types of grasping model training. Bold marking indicates the optimal value under the evaluation criterion

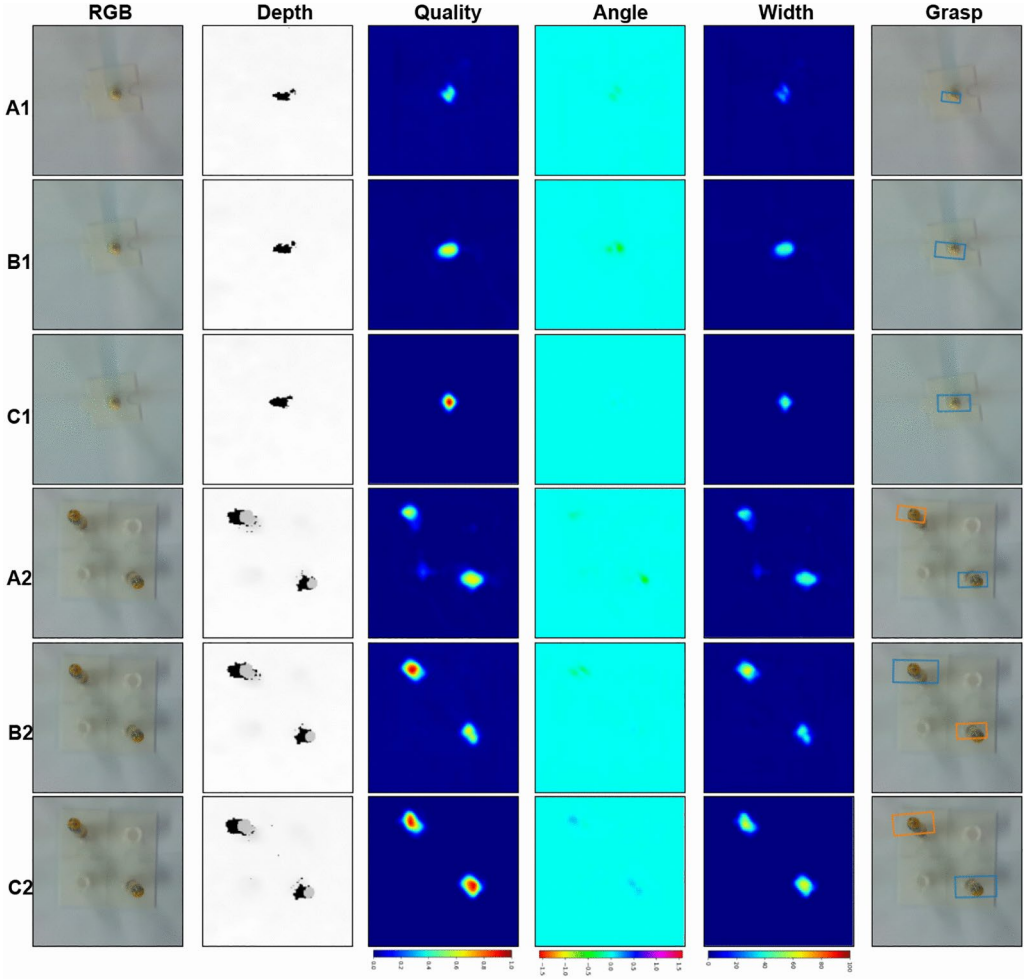

Fig. 11. Single-target and multi-target grasping frame detection results of emulsion explosives. The grasping network takes RGB and Depth images as its inputs, while Quality (ranging from [0, 1]), Angle (ranging from $[\frac{\pi}{2}, \frac{\pi}{2}]$), Width (ranging from [0, 100]), and Grasp are the output generated through inference. GRCNN is utilized for A1 and A2, CBAM-GRCNN for B1 and B2, and SimAM-GRCNN for C1 and C2.

optimizing these critical parameters, comprehensive experiments are conducted to identify the environments most favorable for the robotic execution of explosive grasping and filling tasks.

Grasping detection-based explosive grasping strategy

The study conducts the emulsion explosive grasping experiment using SimAM-GRCNN grasping network, as depicted in Fig. 12. The robotic arm, equipped with a camera, positions itself approximately 30cm above the emulsion explosive to prepare for grasping. It captures an RGB-D image from this viewpoint, which serves as input for the grasping detection process. The SimAM-GRCNN grasping detection system is then invoked, providing detection outcomes. The grasp box with the highest quality score Q is selected and fed into the robot's grasping model, which determines the grasp location in the robot's coordinate system. The robotic arm initiates the grasping maneuver, moving to the designated position, and the gripper closes to seize the explosive. The robot then lifts its end-effector to grasp the explosive and moves it away from the initial placement site, thereby completing the grasping action.

In this section, the study attempts to grasp four explosives of varying colors and sizes, each 20 times, resulting in a total of 80 grasping trials. The outcomes are summarized in Table 7. A grasp is deemed successful when the center line of the emulsion explosive is within ±5 mm of the gripper's central axis. Across all four types of emulsion explosives, the success rate for the 80 grasps is 87.5%. Unsuccessful grasps can be attributed to errors in grasping detection or measurement. Grasping detection errors stem from inaccuracies in the grasp point location or the shape of the grasp bounding box, while grasp measurement errors arise from imprecise distance transformations in the robot's grasping model. On average, each grasping attempt, including SimAM-GRCNN grasping detection and data transmission delays, takes 35 s. The robotic grasping strategy based on grasping detection performs exceptionally well in successfully grasping emulsion explosives of different sizes and colors.

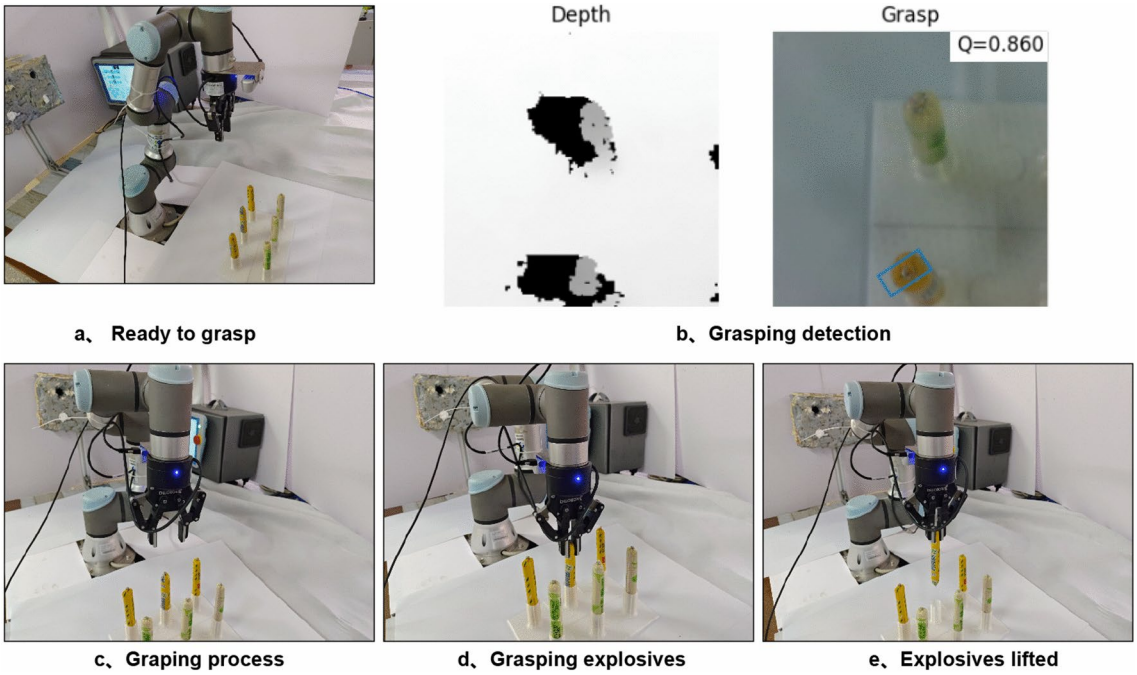

Fig. 12. Explosives grasping strategy based on grasping detection.

Color	Radius (mm)	Number of successful grasping	Number of grasping detection errors	Number of gripping measurement errors
Yellow	9	19	0	1
Yellow	7	18	1	1
Off-white	9	17	1	2
Off-white	7	16	2	2

Table 7. Emulsion explosives grasping strategy results.

Object detection-based blast holes filling strategy

The study conducts an experiment on blast hole filling, which is based on object detection, to validate the reliability of the YOLOv8 blast holes detection model for explosive filling. This experiment is performed on the emulsion explosive grasping and filling test platform. Combining the Evaluation on blast holes dataset for assessment, all subsequent experiments employ the blast holes surface model in silver-gray. The strategy for target detection-based blast holes filling is depicted in Fig. 13. Initially, the robotic arm, equipped with a camera, positions itself approximately 30cm in front of the blast holes at the working face, preparing for the filling process. The camera captures an RGB image of the blast hole surface, which serves as input for the blast holes detection process. The YOLOv8 object detection model is invoked by the robot system, providing the detection results. The location with the highest confidence score is chosen and fed into the robot's grasping model, which determines the filling position in the robot's coordinate system. Subsequently, the robotic arm proceeds to execute the filling process upon receiving the filling coordinates, moving to the designated grasping position. The robot then opens its gripper to deposit the explosive to the specified depth. Finally, the robot retracts its end-effector, completing the explosive filling action, thereby realizing the target detection-based blast holes filling strategy.

The study conducts 20 filling attempts for each of two distinct blast hole sizes, with the outcomes detailed in Table 8. A filling is considered successful when the emulsion explosive reaches a depth within ±5 mm of the target and is centered within ±5 mm of the blast hole center. Specifically, 9 mm radius emulsion explosives are filled into 14 mm diameter blast holes, and 7 mm radius explosives are used for 12 mm diameter holes. Before filling, the gripper's center is aligned to precisely grasp the emulsion explosive. For the 40 filling attempts across both blast hole sizes, an 85% success rate is achieved. Failures in blast holes detection and filling measurement lead to unsuccessful grasps. Misidentified grasp points lead to blast holes detection errors, and inaccuracies in the robot's grasping model transformation distance result in filling measurement mistakes. On average, each filling attempt takes 30 s, including time for YOLOv8 blast holes detection and data transmission for the robot's grasping model transformations. The target detection-based blast hole filling strategy implemented in the robotic filling system demonstrates satisfactory performance across different blast hole sizes.

Having validated the efficacy of the robotic grasping model through cross-validation of the proposed emulsion explosive grasping strategy based on grasping detection and blast hole filling strategy based on object detection, we can directly transplant the robotic framework from the emulsion explosive grasping and filling test

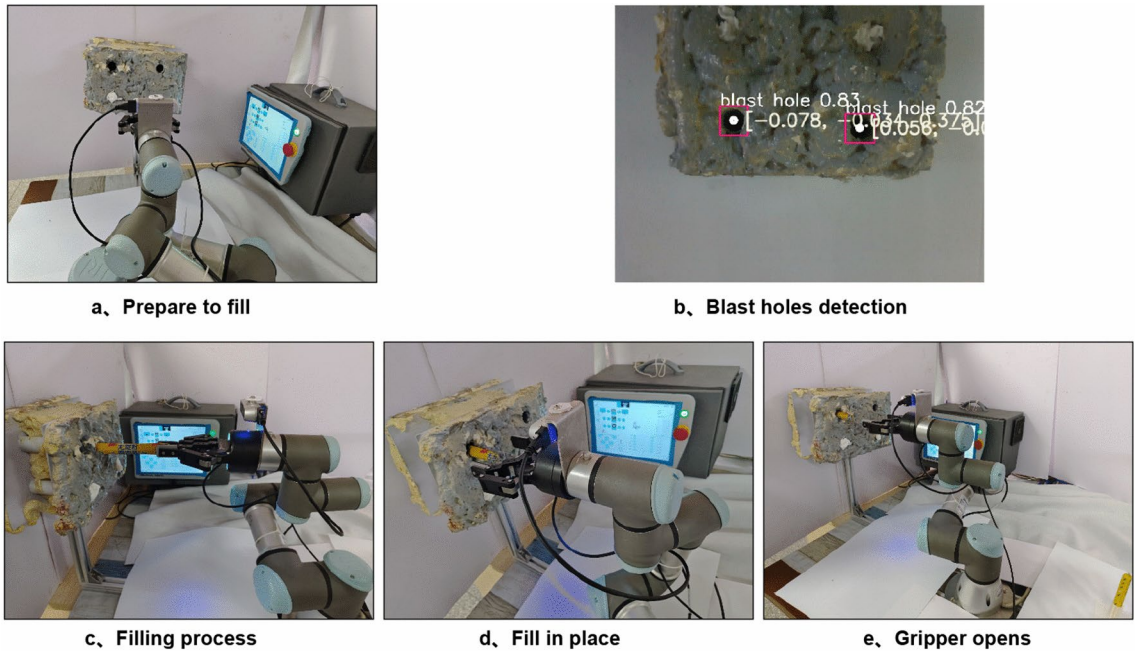

Fig. 13. Blast holes filling strategy based on object detection.

Radius (mm)	Number of successful filling	Number of filling detection errors	Number of filling measurement errors
14	18	0	1
12	16	2	2

Table 8. Blast holes filling strategy results.

platform to a simulated tunnel environment setup for emulsion explosive grasping and filling. This enables us to proceed with experiments based on integrated robotic grasping and filling operations. Figure 14 shows how the UR3 robot grasps multiple emulsion explosives presented in Fig. 4, which can also be referred to as the motion process of the UR3.

The A1 robot system is initialized at the P_home position with the gripper in an open state. Position information for the blast hole is obtained by moving the robot A2 to the front of the blast hole at position P_1. The A3 robot moves to the top of the emulsion explosive at position P_2 to obtain information about the emulsion explosive below. The B1 robot moves to the top of the first emulsion explosive at position $(P_{pick\_i}, 0, 0, 0.2)$ and opens the gripper to an appropriate size. The B2 robot moves to the optimal grasping position P_{pick_i}, closes the gripper to grasp the emulsion explosive. The B3 robot vertically retrieves the emulsion explosive at position $(P_{pick\_i}, 0, 0, 0.2)$. The C1 robot reaches the front of the matching blast hole at position $(P_{fill\_i}, 0, -0.2, 0)$. The C2 robot horizontally fills the emulsion explosive to position P_{fill_i} and opens the gripper. The C3 robot returns to the front of the blast hole at position P_1. To view the complete experimental demonstration of explosive grasping and borehole filling, please click the video link https://www.bilibili.com/video/BV18w411Y7XF to proceed.

Result analysis

In this study, the UR3 robot arm is utilized for experimental verification. The distributed deployment of the YOLOv8 blast holes detection network and SimAM-GRCNN grasping network is integrated with the robot arm motion control mode. Furthermore, these models are deployed on the Jetson Xavier NX embedded development board. A variety of emulsified explosives with different radii, colors, and lighting conditions are selected for grasping and filling experiments. The experimental results obtained at room temperature of 25° are presented in Table 7. Hand-eye calibration is positively correlated with the success rate of grasping; when the center line of the emulsified explosives is judged to be within 5 mm of the gripper's center line, it is considered a successful grasping. The filling accuracy is also positively correlated with the detection network and gripper control; successful filling is determined by meeting the target depth within $\pm$ 2 mm and a range of $\pm$ 5 mm from the center of the blast holes.

The consolidated results in Table 9 indicate that yellow emulsion explosives exhibit higher tractability in detection and grasping compared to their off-white counterparts. Concerning dimensions, emulsion explosives with a 7 mm diameter prove more facile to grasp than those measuring 9 mm. Optimal conditions for both grasping and filling of emulsion explosives are observed when ambient illumination intensity ranges between

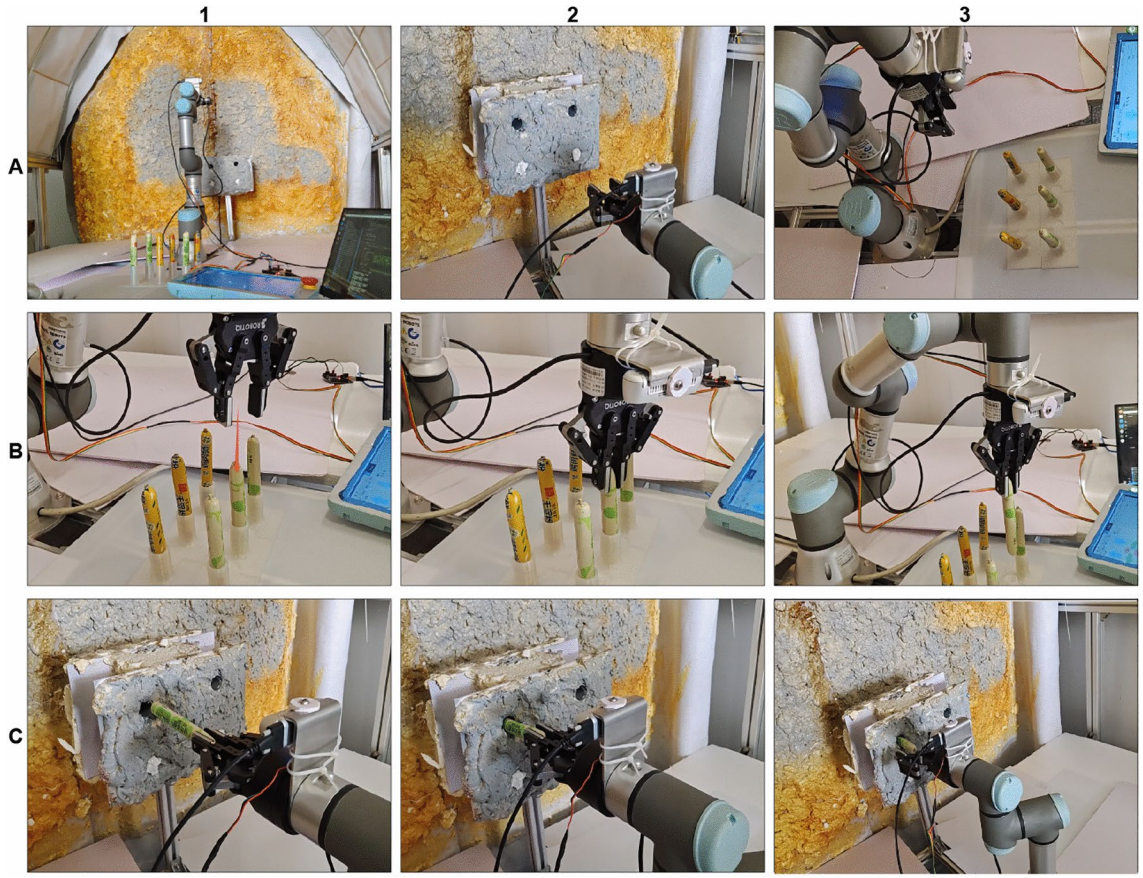

Fig. 14. Schematic diagram of UR3 emulsion explosive grasping.

Color	Radius (mm)	Light intensity	Number	Pick success (%)	Fill success (%)	Average time (s)
Yellow	9	$60 \sim 90$	40	90.0	85.0	117.32
Yellow	9	$90 \sim 120$	40	92.5	87.5	118.63
Yellow	9	$120 \sim 150$	40	87.5	85.0	118.26
Yellow	7	$60 \sim 90$	40	90.0	82.5	117.17
Yellow	7	$90 \sim 120$	40	90.0	85.0	118.85
Yellow	7	$120 \sim 150$	40	87.5	82.5	118.37
Off-white	9	$60 \sim 90$	40	87.5	77.5	120.24
Off-white	9	$90 \sim 120$	40	87.5	80.0	121.57
Off-white	9	$120 \sim 150$	40	85.0	75.0	120.92
Off-white	7	$60 \sim 90$	40	85.0	70.0	120.45
Off-white	7	$90 \sim 120$	40	82.5	72.5	122.72
Off-white	7	$120 \sim 150$	40	80.0	67.5	121.52
Average	8	$90 \sim 120$	40	87.1	79.2	119.67

Table 9. Summary of the grasping effect of different types of emulsion explosives.

90 and 120 Lux. Statistically, the mean success rate for grasping emulsion explosives stands at 87.1%, whereas the combined grasping and filling success rate reaches 79.2%. The aggregate duration for the complete cycle of grasping and filling averages 119.67 s. In actual filling operations, manual or rudimentary equipment typically requires a filling success rate exceeding 75%, despite lacking a standardized benchmark for total time consumption. Our study has achieved the required accuracy in filling. However, the overall time consumption indicates room for improvement. Future refinements in robot motion control can decrease time costs during the grasping and filling phases. The YOLO-SimAM-GRCNN system, with its high portability, seamlessly integrates into various robotic systems, providing an end-to-end solution for target grasping and filling tasks in robotics. This innovation expands the scope of robots' applicability in real-world scenarios.

For real-world application, (1) To improve safety in mining operations, the system utilizes an enhanced grasping model, i.e., the SimAM-GRCNN, trained on an actual grasping and filling dataset. This ensures that the model can accurately detect and grasp the emulsion explosive charges, minimizing the risk of mishandling. By selecting robots with high repeatable positioning accuracy, the system can ensure that the cumulative errors during movement do not interfere with the precise placement of the explosives. Additionally, setting fixed points for the robot's movements and implementing obstacle avoidance functionalities can help restrict the operational range, further enhancing the safety of the filling process. (2) To boost efficiency, the system takes various factors into account that optimize the filling process. For example, choosing emulsion explosives with a radius that matches the size of the robot's gripper ensures a secure and quick grasp. Maintaining an optimal lighting level between 90 and 120 Lux ensures good visibility, facilitating faster and more accurate handling. Furthermore, increasing the speed of the robot's movements, while maintaining control precision, can significantly reduce the time required for each filling operation, thereby increasing overall productivity.

Conclusions and discussions

We have developed a YOLO-SimAM-GRCNN system designed for robots to intelligently grasp and fill emulsion explosives in tunnel blasting scenarios. The system consists of two modules: an inference module and a control module. The inference module incorporates a YOLOv8-based blast hole position detection network and a SimAM-GRCNN-based explosive grasping network. The control module plans and executes the robot's motion control based on the detected blast hole positions and emulsion explosive grasping poses to achieve symmetrical grasping and filling operations. We have designed an improved SimAM-GRCNN grasping model, integrated with an eye-in-hand calibration method to meet the robot's end-to-end grasping requirements. The model has been evaluated on three grasping datasets and a self-built emulsion explosive dataset, showing superior performance in single-object and multi-object grasping compared to another advanced model. By combining the UR3 robotic arm with the Jetson Xavier NX development board, we have successfully deployed the emulsion explosive grasp-and-fill system. The system achieves blast holes detection, emulsion explosive grasping, and filling tasks with an average time of 119.67 s for a complete process. The success rates for grasping emulsion explosives and filling are 87.1% and 79.2%, respectively.

In three-dimensional space, the position and orientation of an object are commonly represented using six degrees of freedom. However, the study employs a four-dimensional representation for robotic grasping, which encompasses planar coordinates and an Euler angle about the z-axis. This approach does not fully capture the complete three-dimensional pose of the object, thereby presenting limitations such as the inability to achieve three-dimensional grasps on the side of objects or to grasp stacked items effectively.

This study has developed a YOLO-SimAM-GRCNN system that enables robotic emulsion explosives grasping and blast holes filling tasks within simulated tunnel environments. To enhance the robot's adaptability in real-world mobile applications, the system can be expanded into a mobile platform, incorporating odometry, LiDAR, and interfacing with the ROS platform. By performing SLAM (Simultaneous Localization and Mapping) within the operational environment, in conjunction with the end-to-end grasping and filling functionalities presented in this work, the system will be well-prepared for practical robotic applications in actual tunnel conditions.

Received: 26 October 2023; Accepted: 18 October 2024
Published online: 18 November 2024

References

1. Davey, R. Smart mining: The benefits of developing digital mines. *AZoMining* (2023).
2. Lenz, I., Lee, H. & Saxena, A. Deep learning for detecting robotic grasps. *Int. J.Robot. Res.* **34**, 705–724 (2013).
3. Zhang, H. et al. A real-time robotic grasping approach with oriented anchor box. *IEEE Trans. Syst. Man Cybern. Syst.* **51**, 3014–3025 (2018).
4. Patten, T., Park, K. & Vincze, M. Dgcm-net: Dense geometrical correspondence matching network for incremental experience-based robotic grasping. *Front. Robot. AI* **7**, 120 (2020).
5. Ma, L. et al. A method of grasping detection for kiwifruit harvesting robot based on deep learning. *Agronomy* **12**, 3096 (2022).
6. LeCun, Y., Bottou, L., Bengio, Y. & Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **86**, 2278–2324 (1998).
7. Glenn, J., Ayush, C., Alex, S. & et al. ultralytics/yolov5: v7.0 - yolov5 sota realtime instance segmentation (v7.0). Zenodo (2022).
8. Terven, J. R. & Esparza, D. M. C. A comprehensive review of yolo: From yolov1 to yolov8 and beyond. arXiv: 2304.00501 (2023).
9. Wang, C.-Y., Bochkovskiy, A. & Liao, H.-Y. M. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, **1**, 7464–7475 (2022).
10. Jocher, G., Chaurasia, A. & Qiu, J. Yolo by ultralytics. Accessed 30 Feb 2023; https://github.com/ultralytics/ultralytics (2023).
11. Yang, L., Zhang, R.-Y., Li, L. & Xie, X. Simam: A simple, parameter-free attention module for convolutional neural networks. In: *International Conference on Machine Learning*, 11863–11874 (2021).
12. Kumra, S., Joshi, S. & Sahin, F. Antipodal robotic grasping using generative residual convolutional neural network. In: *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, **1**, 9626–9633 (2019).
13. Maitin-Shepard, J. B., Cusumano-Towner, M. F., Lei, J. & Abbeel, P. Cloth grasp point detection based on multiple-view geometric cues with application to robotic towel folding. In: *2010 IEEE International Conference on Robotics and Automation*, **1**, 2308–2315 (2010).
14. Lv, Y., Yuan, R. & Song, G. Multivariate empirical mode decomposition and its application to fault diagnosis of rolling bearing. *Mech. Syst. Signal Process.* **81**, 219–234 (2016).
15. Hu, J., Li, Q. & Bai, Q. Research on robot grasping based on deep learning for real-life scenarios. *Micromachines* **14**, 1392 (2023).
16. Song, Q. et al. Object detection method for grasping robot based on improved yolov5. *Micromachines* **12**, 1273 (2021).
17. Alaaudeen, K., Selvarajan, S., Manoharan, H. & Jhaveri, R. H. Intelligent robotics harvesting system process for fruits grasping prediction. *Sci. Rep.* **14**, 2820 (2024).
18. Wanzhi, Z. Algorithm for Image Recognition of Rock Tunnel Blast Holes and Optimization of Smooth Surface Blasting Parameters. Ph.D. thesis, Shandong University (2019).

19. Ye, Z. Research on Related Technologies of Blast Hole Recognition and Feasible Area Planning for Intelligent Explosive Loading Robot. Ph.D. thesis, Liaoning University of Science and Technology (2020).
20. Zhongwen, Y. et al. Research on lightweight intelligent detection method for blast holes based on deep learning. *J. Coal Sci.* **1**, 1–12 (2023).
21. Li, Z. & Ren, D. Improved yolov8 based small object detection for intelligent robotic arm in complex environments. In: *2024 IEEE 13th Data Driven Control and Learning Systems Conference (DDCLS)*, 1124–1130. https://doi.org/10.1109/DDCLS61622.2024.10606904 (2024).
22. Zhong, X., Chen, Y., Luo, J., Shi, C. & Hu, H. A novel grasp detection algorithm with multi-target semantic segmentation for a robot to manipulate cluttered objects. *Machines* **12**, 506 (2024).
23. Jin, Y. et al. Target localization and grasping of NAO robot based on yolov8 network and monocular ranging. *Electronics*[SPACE]https://doi.org/10.3390/electronics12183981 (2023).
24. Kumar, A. & Behera, L. High-speed detector for low-powered devices in aerial grasping. *IEEE Robotics and Automation Letters* (2024).
25. Kolin, N. & Chebotareva, E. A comparative analysis of object detection methods for robotic grasping. In: *2024 International Conference on Artificial Life and Robotics (ICAROB2024)* (2024).
26. Yan, B., Liu, Y. & Yan, W. A novel fusion perception algorithm of tree branch/trunk and apple for harvesting robot based on improved yolov8s. *Agronomy*[SPACE]https://doi.org/10.3390/agronomy14091895 (2024).
27. Bicchi, A. & Kumar, V. R. Robotic grasping and contact: a review. Proceedings 2000 ICRA. Millennium Conference. In: *IEEE International Conference on Robotics and Automation*. Symposia Proceedings (Cat. No.00CH37065) **1**, 348–353 (2000).
28. Atique, M. M. U. & Francis, J. T. Mirror neurons are modulated by grip force and reward expectation in the sensorimotor cortices (s1, m1, pmd, pmv). *Sci. Rep.* **11**, 15959 (2021).
29. Satish, V., Mahler, J. & Goldberg, K. On-policy dataset synthesis for learning robot grasping policies using fully convolutional deep networks. *IEEE Robot. Autom. Lett.* **4**, 1357–1364 (2019).
30. Schmidt, P., Vahrenkamp, N., Wächter, M. & Asfour, T. Grasping of unknown objects using deep convolutional neural networks based on depth images. In: *2018 IEEE International Conference on Robotics and Automation (ICRA)*, **1**, 6831–6838 (2018).
31. Zeng, A. et al. Robotic pick-and-place of novel objects in clutter with multi-affordance grasping and cross-domain image matching. *Int. J. Robot. Res.* **41**, 690–705 (2017).
32. Peng, G., Liao, J., Guan, S., Yang, J. & Li, X. A pushing-grasping collaborative method based on deep q-network algorithm in dual viewpoints. *Sci. Rep.* **12**, 3927 (2021).
33. Mousavian, A., Eppner, C. & Fox, D. 6-dof graspnet: Variational grasp generation for object manipulation. In: *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, **10**, 2901–2910 (2019).
34. Murali, A., Mousavian, A., Eppner, C., Paxton, C. & Fox, D. 6-dof grasping for target-driven object manipulation in clutter. In: *2020 IEEE International Conference on Robotics and Automation (ICRA)*, **1**, 6232–6238 (2019).
35. Kumra, S., Joshi, S. & Sahin, F. Gr-convnet v2: A real-time multi-grasp detection network for robotic grasping. *Sensors (Basel Switzerland)* **22**, 6208–6233 (2022).
36. Ge, J., Shi, J., Zhou, Z., Wang, Z. & Qian, Q. A grasping posture estimation method based on 3d detection network. *Comput. Electr. Eng.* **100**, 107896 (2022).
37. Bin, Z., Chengdong, W. & Xuejiao, Z. e. a. Mechanical arm object grasping network technology based on attention mechanism. Journal of Jilin University (Engineering Edition) 1–9 (2023).
38. Redmon, J. & Angelova, A. Real-time grasp detection using convolutional neural networks. In: *2015 IEEE International Conference on Robotics and Automation (ICRA)*, **1**, 1316–1322 (2014).
39. Zhou, X. et al. Fully convolutional grasp detection network with oriented anchor box. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, **1**, 7223–7230 (2018).
40. Morrison, D., Corke, P. & Leitner, J. Learning robust, real-time, reactive robotic grasping. *Int. J. Robot. Res.* **39**, 183–201 (2019).
41. Yu, S., Zhai, D., Xia, Y., Wu, H. & Liao, J.-J. Se-resunet: A novel robotic grasp detection method. *IEEE Robot. Autom. Lett.* **1**, 5238–5245 (2022).
42. Guozhong, L. & Liangwen, S. Technical appraisal of bcj-1 underground small and medium diameter bulk emulsion explosive charging vehicle led by the national defense science and technology commission. *Nonferrous Metals* **01**, 84–84 (2002).
43. Chunsheng, H. et al. Research and application of palletizing robots. *Comput. Eng. Appl.* **58**, 57–77 (2022).
44. Hongpeng, C., Bing, G. & Xin, L. Motion characteristics analysis of underground intelligent explosive charging vehicle. *Mining Metall.* **24**, 4 (2015).
45. Yuan, R., Lv, Y., Wang, T., Li, S. & Li, H. Looseness monitoring of multiple m1 bolt joints using multivariate intrinsic multiscale entropy analysis and Lorentz signal-enhanced piezoelectric active sensing. *Struct. Health Monit. Int. J.* **21**, 2851–2873 (2022).
46. Zhang, Q., Yuan, R., Lv, Y., Li, Z. & Wu, H.-Y. Multivariate dynamic mode decomposition and its application to bearing fault diagnosis. *IEEE Sensors J.* **23**, 7514–7524 (2023).
47. Hara, K., Vemulapalli, R. & Chellappa, R. Designing deep convolutional neural networks for continuous object orientation estimation. ArXiv:abs/1702.01499 (2017).
48. Pozzi, L. et al. Grasping learning, optimization, and knowledge transfer in the robotics field. *Sci. Rep.* **12**, 4481 (2022).
49. Huijun, J. Research on Key Technologies of Industrial Robot Gripping System for Digital Twin. Ph.D. thesis, North University of China (2022).
50. Yao, X. Modeling of Working Environment and Coordinated Motion Planning for Mobile Manipulation Robots. Ph.D. thesis, Dalian University of Technology (2021).
51. Jingjing, M. et al. Design of intelligent packaging system for automated production line based on plc and industrial robots. *Manuf. Technol. Mach. Tool* **11**, 63–67 (2021).
52. Depierre, A., Dellandréa, E. & Chen, L. Jacquard: A large scale dataset for robotic grasp detection. In: *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, **1**, 3511–3516 (2018).
53. Jiang, Y., Moseson, S. & Saxena, A. Efficient grasping from rgbd images: Learning using a new rectangle representation. In: *2011 IEEE International Conference on Robotics and Automation*, **2**, 3304–3311 (2011).
54. Kumra, S. & Kanan, C. Robotic grasp detection using deep convolutional neural networks. In: *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, **1**, 769–776 (2016).
55. Asif, U., Tang, J. & Harrer, S. Graspnet: An efficient convolutional neural network for real-time grasp detection for low-powered devices. In: *International Joint Conference on Artificial Intelligence*, 4875–4882 (2018).

Acknowledgements

This research is supported by Hubei Provincial Natural Science Foundation of China (General Program: No.2023AFB993, Youth Program: No.2023AFB028, Innovation Development Joint Key Program: No.2023AFD001), State Key Laboratory of Precision Blasting and Hubei Key Laboratory of Blasting Engineering, Jianghan University (No.PBSKL2022302, No.PBSKL2023B3), Hubei Key Laboratory of Industrial Fume and Dust Pollution Control, Jianghan University (No.HBIK2023-03), and the Research Fund of Jianghan University (Grant No. 2023KJZX34), which are greatly appreciated.

Author contributions

J.Y. performed the methodology, supervision, and funding acquisition. P.L. was mainly responsible for the software, formal analysis, wrote the manuscript, and made edits. J.G. performed the methodology, investigation, supervision and funding acquisition. R.Y. accounted for the resources, visualization and funding acquisition. J.W. was mainly responsible for the software and formal analysis. All authors provided critical feedback and contributed to the final manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-77034-0.

Correspondence and requests for materials should be addressed to J.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.