Article citation info:
Zhou W, Xu Y, Huang Y. Design of automatic target recognition and detection system for sensors based on improved machine vision control. Diagnostyka. 2025;26(2):2025206. https://doi.org/10.29354/diag/203646.

DIAGNOSTYKA, 2025, Vol. 26, No. 2

e-ISSN 2449-5220
DOI: 10.29354/diag/203646

DESIGN OF AUTOMATIC TARGET RECOGNITION AND DETECTION SYSTEM FOR SENSORS BASED ON IMPROVED MACHINE VISION CONTROL

Wei ZHOU *, Ye XU, Yue HUANG

Wuxi Institute of Technology, China
* Corresponding author, e-mail: zhou-wei163@hotmail.com

Abstracts

The aim of this study is to design and implement a sensor-based automatic target recognition and detection system with machine vision control. The system achieves high-precision detection of targets in complex environments by integrating multiple sensors, including industrial-grade color and infrared cameras, VelodyneHDL-64E lidar, ultrasonic arrays, and high-quality IMU devices. Through multi-sensor data fusion, preprocessing techniques, and feature extraction methods, the experimental results show that the system is able to achieve high-precision target detection in different scenarios. FasterR-CNN and its improved version of the model perform well in the experiments, especially after the introduction of the feature pyramid network (FPN) and the attention mechanism, which significantly improves the detection rate and the overall performance of the small targets. Experimental results show that the multi-sensor fusion system significantly improves the performance in target detection, with the accuracy of RGB cameras increasing from 85% to 92% and the recall rate increasing from 78% to 88%. After introducing the feature pyramid network (FPN) and attention mechanism, the detection accuracy of the Faster R-CNN model for small targets increased from 70% to 75%. Although the processing speed decreased slightly (from 20fps to 15fps), the overall detection accuracy and robustness were significantly enhanced. In addition, the model pruning technology increased the processing speed to 12fps while maintaining high accuracy, which is suitable for real-time applications. The model pruning technique successfully realizes the lightweighting of the model while maintaining high detection accuracy, which provides the possibility of real-time target detection for embedded devices.

Keywords: machine vision, vision control, automatic target recognition and monitoring

1. INTRODUCTION

In the context of today's booming information technology and artificial intelligence technology, machine vision technology has become the core force to promote cutting-edge innovation in science and technology, especially in many fields such as industrialized production, intelligent security systems, and automated driving, and its wide range of applications are subconsciously rewriting the mode of human life and work. On the intelligent manufacturing assembly line, through the integration of advanced machine vision systems, product details can be precise quality inspection and real-time production tracking, which significantly improves the working efficiency of the production line and product quality [1]. In the intelligent security system, the application of machine vision technology is embodied in face recognition and behavior analysis, through real-time monitoring and intelligent analysis, it greatly strengthens the protection effectiveness of the security system, and effectively prevents and combats criminal behavior [2]. In the field of automatic driving technology, the vehicle-mounted machine vision system is indispensable, capable of real-time, accurate identification of road pedestrians, vehicles, traffic signs and other key information, for the safety and reliability of automatic driving to build a solid line of defense [3]. Sensor automatic target recognition and detection system as the cornerstone component of machine vision technology, its performance directly determines the effectiveness of the whole system. The processing of machine vision is shown in Figure 1, mainly including data acquisition, transmission and processing.

Currently, the research and development of sensor-based automatic target recognition and detection technology presents a blossoming situation. On the one hand, a series of deep learning-based target detection networks represented by YOLO (YouOnlyLookOnce) and FasterR-CNN have achieved a disruptive breakthrough in the field of image recognition by virtue of their strong advantages in the level of feature learning and abstraction, which have substantially improved the speed and accuracy of target detection [4].

Received 2024-07-27; Accepted 2025-04-02; Available online 2025-04-04

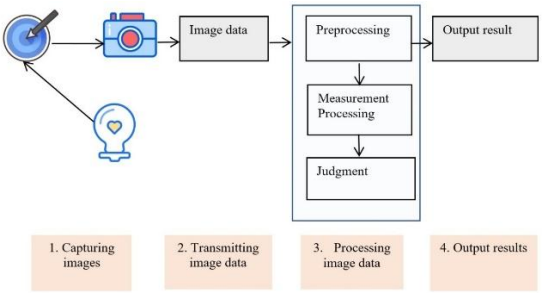

Fig. 1. Machine vision processing

However, deep learning methods also face some challenges that cannot be ignored, such as the high dependence on a large amount of training data, the high consumption of computational resources, and the decline of recognition accuracy in special environments such as complex lighting and target occlusion [5]. On the other hand, traditional image processing techniques, such as template matching, edge detection, corner detection, etc., have survived the ages but still maintain their unique value in specific application scenarios. These methods still have a place in some specific tasks by virtue of the simplicity of the algorithms and the fast real-time response time. However, these traditional methods tend to rely too much on human-designed features and are relatively weakly robust to target geometric transformations, especially when dealing with complex backgrounds, diverse target types, and variations in target size and attitude, their performance tends to be limited [6].

The current sensor-based automatic target recognition and detection system faces severe challenges in terms of recognition accuracy, real-time response capability, anti-interference capability and environmental adaptability. Especially in extreme situations such as drastic changes in lighting conditions, encountering bad weather, and complex variations in target size and attitude, the performance of the system may experience significant slippage, which is a major challenge that needs to be urgently solved in the current research field [7]. Therefore, how to develop the sensor automatic target recognition and detection technology with high accuracy, real-time response, strong anti-interference ability and good environmental adaptability has become a focus of researchers and industry to focus on and overcome.

This study aims to improve and refine the machine vision control strategy through in-depth research and technological innovation in response to the problems of the existing sensor-based automatic target recognition and detection system, with a view to comprehensively enhancing the comprehensive performance of the system. The specific research includes but is not limited to the following aspects:

(1) propose novel visual information processing algorithms, which are expected to improve the adaptability to complex environments and multi-targets while reducing the computational complexity, so as to improve the accuracy and real-time performance of target recognition.
(2) Design an optimized sensor fusion mechanism to achieve complementary advantages by integrating information from different types of sensors to enhance the system's anti-interference capability and stability.

Overall, this research not only focuses on the breakthrough of existing technical bottlenecks, but also actively explores and constructs new methods of sensor-based automatic target recognition and detection that are more efficient, reliable and widely adaptable to provide strong support for promoting the in-depth application of machine vision technology in an intelligent society.

In the field of technical diagnosis, the automatic target recognition and detection system based on improved machine vision control can provide strong support through deep learning models such as Faster R-CNN. The system first uses multi-sensor fusion technology to collect data from different sources (such as visible light cameras, infrared cameras, and LiDAR) to obtain comprehensive information about the object or environment to be detected. After denoising and calibrating these data in the preprocessing stage, the feature pyramid network (FPN) ensures that the cross-scale target characteristics are effectively captured, while the attention mechanism helps focus on key details and reduce the influence of irrelevant background. Through the trained Faster R-CNN model, high-precision positioning and classification of specific fault modes or abnormal states can be achieved. For example, in industrial maintenance, the system can quickly identify tiny cracks or other potential defects on equipment; in medical image analysis, it can assist doctors in discovering the location of early lesions. This method, which combines advanced visual algorithms with intelligent perception technology, not only improves diagnostic efficiency, but also enhances the reliability of the results, providing a strong basis for maintenance decisions.

2. THEORETICAL FOUNDATION AND KEY TECHNOLOGY

2.1. Basic theory of machine vision

The fundamental body of theory of machine vision, especially elaborated exhaustively in the study of [8], constitutes a complete methodology, which guides all the key steps from image capture up to the precise recognition and localization of the target. The image acquisition phase, deeply rooted in the principles of optical imaging, which is explained in depth in the study of [9], points out that by using advanced optoelectronic sensor technologies, such as charge-coupled devices (CCDs) or complementary metal-oxide-semiconductor (CMOS) sensors, complex three-dimensional solid spatial information can be efficiently mapped onto a two-dimensional image

plane, realizing the conversion of the physical world into a digital image.

Moving to the preprocessing stage, in order to enhance the quality of the image for subsequent analysis, researchers have widely applied a variety of mathematical models and algorithms. As discussed in [10], these methods include, but are not limited to, the use of filtering techniques to eliminate image noise, such as mean filtering, median filtering, adaptive filtering, etc., as well as the use of contrast stretching, histogram equalization, etc. to enhance the overall or local contrast of the image, thereby improving the clarity and visibility of the image.

Feature extraction theory occupies a central position in machine vision, and advances in this field are reflected in the research of [11], who explored how contour and boundary information in images can be captured by edge detection techniques, and corner detection techniques are used to identify those key points in an image that represent significant geometric changes. In addition, advanced feature descriptors such as SIFT, SURF, ORB, etc., which are notable for their invariance to changes in scale, rotation, and even illumination, are effective in capturing and encoding key structural features of an image, which are crucial for subsequent target identification and matching.

Facing the complexities encountered in practical applications, such as changing lighting conditions [12], partial or complete occlusion [13], scale scaling and viewpoint transformation [14], researchers have continuously proposed and optimized solutions. For example, [15] made a breakthrough in the field of light invariant features and developed a feature representation that can resist light interference; the multi-view learning strategy proposed by [16], on the other hand, helps the system to accurately recognize the same target under different viewing perspectives; in recent years, with the rise of deep learning technology, [17] showed in their study how to fuse multi-scale features in deep neural networks, and this strategy successfully copes with the challenges posed by changes in target size and viewpoints, and improves the accuracy and stability of recognition and localization.

2.2. Sensor technology and characterization

A variety of sensors play an indispensable role in machine vision systems [18], each playing a unique advantage, and together provide strong support for the efficient operation of the system. At the image acquisition level, charge-coupled device (CCD) and complementary metal-oxide-semiconductor (CMOS) cameras are suitable for different application scenarios due to their respective technical characteristics; CCD cameras dominate high-precision image acquisition with their excellent photosensitivity and signal-to-noise ratio, which are especially suitable for scientific research and professional-grade applications [19]; whereas, CMOS cameras, with their low-cost, low-power and high-speed readout characteristics, have been widely used in high-volume consumer-grade applications. readout, CMOS cameras have shown competitiveness in the high-volume consumer-grade market and cost-sensitive projects [20].

The rapid development of LiDAR (laser radar) technology [21] has made it a key component of 3D spatial environment modeling and obstacle detection, especially in the field of self-driving cars and drones. Meanwhile, infrared thermal cameras (e.g., a product of FlirSystemsInc.,2018) utilize infrared energy radiated from the surface of an object for imaging, and are able to provide unique visual information in poor visual conditions, such as low-light, smoggy, or nighttime conditions. Ultrasonic sensors, on the other hand, specialize in short-range object detection and ranging, and are commonly used in scenarios such as indoor navigation, obstacle avoidance, and liquid level monitoring.

Sensor fusion technology [22] is an important strategy to overcome the limitations of a single sensor by integrating data from multiple sensors and utilizing their complementarities to improve the reliability and accuracy of the system. For example, camera and radar data synergy [23] is particularly important in self-driving vehicles, where the camera provides rich visual information while the radar compensates for the shortcomings in bad weather and long range detection.

However, sensor fusion is not an easy task, and the technical challenges involved include how to achieve accurate synchronization of multi-sensor data [24], and selecting and optimizing fusion algorithms suitable for specific application scenarios [25]. In addition, many studies have explored topics such as the design of fusion architectures, the handling of sensor uncertainty [26], and target tracking in multisensor environments, with a view to further enhancing the effectiveness and adaptability of machine vision systems.

2.3. Target recognition and detection algorithms

The research on the application of deep learning in the field of target recognition and detection has experienced a transition from initial exploration to deep penetration. The framework in the field of target recognition and detection is shown in Figure 2. Initially, classical target detection algorithms such as Viola-Jones (Viola & Jones, 2001) utilized cascade classifiers to achieve face detection, laying the foundation for subsequent algorithm development. With the development of deep learning technology, the R-CNN (Region-basedConvolutionalNeuralNetworks) family has gradually emerged [27]. FastR-CNN reduces the computational overhead by sharing convolutional layers, while [28] further introduces the Region Proposition Network (RPN), which greatly improves the detection speed and accuracy.

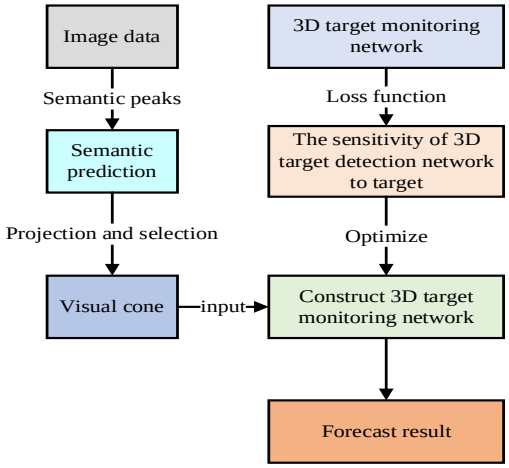

Fig. 2. Target recognition and detection

The specific flowchart is shown in Fig. 2. The YOLO (YouOnlyLookOnce) family of algorithms [29] pioneered a one-step detection paradigm, which realizes real-time target detection by predicting the bounding box and category probabilities in the whole image at once. Similar to this is SSD [30], which directly predicts the location and category of an object by setting multiple anchor frames on different feature layers. MaskR-CN introduces the concept of instance segmentation, which not only detects the location of a target, but also accurately segments the pixel-level contour of each target. In addition, [31] solved the problem of imbalance between foreground and background samples by introducing focalloss, which improved the detection of small targets. Despite the remarkable achievements of deep learning methods in target recognition and detection, they still face challenges when dealing with complex scenarios such as small target detection [32], occlusion target detection [33], and dense target detection. Researchers continue to explore new network structures [33] and loss function optimization strategies (e.g., IoU-smoothL1Loss to improve algorithm performance and generalization ability. Traditional image processing methods are still useful in specific scenarios. For example, HOG (HistogramofOrientedGradients) and SIFT (Scale-InvariantFeatureTransform) [34] still have high recognition accuracies in some specific target types and constrained environments. However, compared to deep learning methods, the performance of these traditional methods is limited when dealing with large-scale datasets, complex backgrounds, and multi-category targets, and thus are more often used as complementary or auxiliary tools in modern target recognition and detection tasks [35].

In summary, research on target recognition and detection algorithms has transitioned from traditional image processing methods to deep learning-led solutions, and new network architectures and optimization strategies continue to emerge to address more complex real-world problems. Future research will continue to focus on improving the accuracy and real-time performance of algorithms, as well as on how to better address the complex challenges of real-world applications system design.

3. SENSOR AUTOMATIC TARGET RECOGNITION AND DETECTION SYSTEM DESIGN

3.1. Multi-sensor integration scheme

When building an automatic target recognition and detection system with sensors based on machine vision control, the first task is to reasonably integrate multiple sensors to form a multimodal, high-precision information acquisition system. The system usually includes, but is not limited to, visible cameras (CV), infrared cameras (IR), LiDAR, ultrasonic sensors, and inertial measurement units (IMUs). These sensors are arranged at different observation levels and angles according to their characteristics and functions to maximize the information acquisition range and accuracy. The simultaneous acquisition and fusion of multi-sensor data aims to overcome the limitations of a single sensor in complex environments [36].

The key to achieving multi-sensor integration lies in ensuring time synchronization and compatibility of data formats and transmission protocols.

3.2. Data preprocessing and feature extraction module design

The data preprocessing stage is crucial to improve the accuracy of subsequent target detection. Raw sensor data first needs to be denoised, for example, image noise can be removed using algorithms such as median filtering or Gaussian filtering: Filtered $Image = Noise\ Reduction(I_{raw})$. Calibration operations are then used to eliminate systematic errors in the sensor itself, such as aberration correction, projection transformation, etc., to ensure the authenticity and accuracy of the data. Data smoothing operations, such as moving average, are used to reduce data fluctuations and improve signal quality.

Different strategies are adopted for feature extraction for different types of sensor data. For image sensors, edge detection operators (e.g., Sobel, Canny, etc.) can be applied to extract edge features: or corner detection algorithms (e.g., Harris corner detection, SIFT feature points, etc.) can be used to search for salient feature points in the image: $K = Corner\ Detection(I_{filtered})$ and further enrich the feature expression by means of texture analysis and color feature extraction. As for point cloud data (e.g., from LIDAR and ultrasonic sensors), the focus is on the geometric features and spatial relationships of the point cloud, such as point cloud clustering, PCA dimensionality reduction,

and feature histograms *PointCloud Features* = *Feature Extraction*(P_{cloud}) . In order to ensure the consistency of the features and facilitate the processing of the deep learning model, it is also necessary to carry out standardization and normalization operations on the extracted features, such as zero-mean normalization (Z-scorenormalization): $X' = \frac{X-\mu}{\sigma}$ where μ is the mean value of the features and σ is the standard deviation of the features [37].

3.3. Deep learning based target detection module design

The deep learning-based target detection module is the core part of the system, which adopts mainstream network architectures such as FasterR-CNN, YOLO and SSD. In the design process, the first step is to select the appropriate network structure according to the actual application scenario and hardware resources, and carry out customized design. For example, in order to solve the problem of target detection with different sizes, a feature pyramid network (FPN) can be introduced to realize multi-scale feature fusion, the basic principle of which is $P_l = \mathrm{FPN}(C_l, C_{l+1}, \ldots, C_L)$. Where P_l is the lth layer pyramid feature map, and C_l to C_L are the deep to shallow convolutional feature maps, respectively [38].

To jointly optimize the classification and localization tasks, a composite loss function is constructed, such as a multi-task loss function, including a classification loss L_{cls} and a bounding box regression loss L_{box}: $L = L_{cls} + \lambda L_{box}$. Among them, λ is the weight factor to balance the two types of losses. The network parameters are optimized by training a large amount of labeled sample data to improve the target recognition rate and localization accuracy of the model in practical applications.

4 AUTOMATIC TARGET RECOGNITION AND DETECTION OF SENSORS BASED ON IMPROVED MACHINE VISION CONTROL

FasterR-CNN, an advanced deep learning target detection algorithm, has achieved breakthroughs in the field of visual perception. The architecture consists of two key components: the RegionProposalNetwork (RPN) and the subsequent RoI pooling layer, which is ultimately succeeded by a full convolutional neural network for category classification and bounding box pinpointing.

4.1. Fasterr-Cnn

1) *Bounding Box Regression for Regional Proposal Networks (RPNs)*

Given an anchor point A_i, its regression target t contains four elements, which are the center coordinate offset and the width and height scaling factor: $t_i = (t_x t_y t_w t_h)$. Where the offset relative to the center coordinate of the anchor box A_i is defined as: $t_x = \frac{x-x_c}{w}$ and $t_y = \frac{y-y_c}{h}$ Here, (x, y) is the center coordinate of the real bounding box and (x_c, y_c) is the center coordinate of the anchor box. The width and height scaling factor is defined as: $t_w = log(\frac{w}{w})$ $t_h = log(\frac{h}{h})$. The regression loss function used by RPN is usually chosen to be SmoothL1Loss, and the regression loss for the ith anchor frame can be expressed as: $L_{reg}(t_i, t_i) = SmoothL1(t_i - t_i)$. The RPN flowchart is specifically shown in Fig. 3.

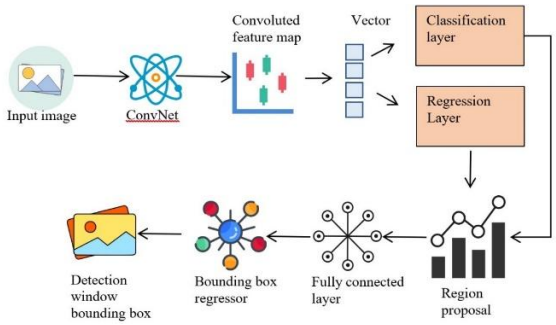

Fig. 3. RPN mode diagram

2) *Binary Classification Losses in Regional Proposal Networks (RPNs)*

For each anchor box A_i, its corresponding binary classification task predicts a probability of belonging to the foreground (containing the target) or the background p_i. Using binary cross entropy as the classification loss function, assuming that there are $K + 1$ categories (where category K is the target category and category K+1 is the background), the classification loss for anchor box A_i is computed as $L_{cls}(p_i, p_i) = -p_i^{log}\,{}_{ii}\,{}^{)log(}\,{}_{i}$ where p_i is the true label corresponding to the anchor box A_i , which takes the value of 1 when the anchor box covers the true bounding box and 0 otherwise [39].

3) *Quantization operations in the roipooling layer*

For the candidate region R, RoIPooling quantizes it into a fixed-size feature map of $H' \times W'$. Assuming that $R = (x_1, y_1, x_2, y_2)$ is the normalized candidate region coordinates (ranging from 0 to 1) and H and W are the height and width of the feature map, respectively, the quantization process can be expressed as follows:

$\hat{R}_{ij} = MaxPool(\boldsymbol{F}(x, y) \mid x \in [(x_1 + \frac{i}{H'})W, (x_2 + \frac{i+1}{H'})W],\ y \in [(y_1 + \frac{j}{W'})H, (y_2 + \frac{j+1}{W'})H])$. Where $\boldsymbol{F}$ is the underlying feature map and $\hat{R}_{ij}$ is the maximum value at position index (i,j) on the quantized feature map [40].

4.2. Improved method based on machine vision control

1) *Characteristic Pyramid Networks*

FeaturePyramidNetwork (FPN) is an innovative architecture for solving the multi-scale target detection problem in target detection tasks. In

traditional deep convolutional neural networks, the spatial resolution of the feature maps decreases as the depth of the network increases, and although higher-level semantic information can be obtained, the detection performance for small-size targets decreases. On the contrary, although the shallow feature maps have higher spatial resolution, the semantic information is coarser, which is not conducive to recognizing target categories. By fusing different levels of feature maps, FPN is able to retain both the semantic information of high-level features and the spatial information of low-level features, resulting in good detection performance on targets of different scales. It is difficult to handle large, medium and small sized targets simultaneously on a single scale feature map, while FPN makes the network flexible to adapt to various sized targets in the detection process by constructing a multi-scale feature pyramid. In traditional deep convolutional neural networks, the spatial resolution of the feature map gradually decreases as the depth of the network increases, and although higher-level semantic information can be obtained, the detection performance for small-sized targets decreases. On the contrary, although the shallow feature maps have higher spatial resolution, the semantic information is coarser, which is not conducive to recognizing target categories. In FPN, by restoring the high-level (low-resolution, containing rich semantic information) feature maps to the same resolution as the low-level feature maps by downsampling (e.g., bilinear interpolation or convolution with a step size of 2). The downsampling formula can be expressed as (assuming here that downsampling is performed using bilinear interpolation): $P_{l'} = \text{Resize}(C_l, size = P_s.shape)$. Next, the downsampled high-level feature map $P_{l'}$ is fused with the unaltered resolution of the low-level feature mapC_s, usually by Element-wiseAddition to achieve feature fusion and information transfer. The formula for the fusion operation$P_{s'} = P_{l'} + C_s$ [40].

2) *Integration of attention mechanisms*

In the target detection task, the attention mechanism can help the model to focus on the target region and ignore the background noise. For example, the SE module (Squeeze-and-ExcitationModule) applied to channel attention first performs a global average pooling of the feature map to obtain a global descriptor for each channel: $z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_{cij}$, where$z_c$ is the global descriptor of the cth channel, x_{cij} is the value of the cth channel of the feature map at the location(i, j), and H and W are the height and width of the feature map, respectively. Next, each channel is assigned a weight value by two fully connected layers and an activation function (e.g., sigmoid or ReLU):

$s_c = \sigma(W_2\delta(W_1 z_c + b_1) + b_2)$ Where, s_c is the attention weight of the cth channel, $W_1 W_2$ is the weight matrix of the fully connected layer, , $b_1 b_2$ are the bias terms,δ is the ReLU activation function, andσ is the sigmoid activation function. Finally, the attention weights are applied to the original feature map to generate a weighted feature map: $y_c = s_c \cdot x_c$. The self-attention mechanism is specifically shown in Fig. 4.

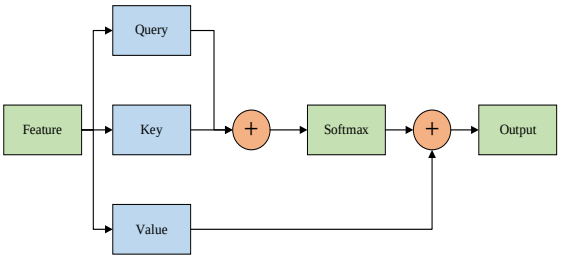

Fig. 4. Self-attention mechanism

3) *Real-time optimization and model lightweighting*

In order to achieve efficient real-time target detection on embedded devices, model optimization is crucial, especially through model pruning techniques to reduce computational complexity and memory usage. Model pruning is essentially a process of slimming down a neural network model by identifying and removing those weight connections that have less impact on the final prediction result, thereby simplifying the model structure and improving the inference efficiency. To implement model pruning, the absolute value of the weight matrix W in each layer of the network is first calculated, and a threshold value θ is set, and weights with an absolute value less than θ are regarded as those that contribute less to the prediction and are removed. The formula is expressed as follows:

$W_{pruned} = W \odot M$, where W_{pruned} is the weight matrix after pruning,$\odot$ denotes the element-by-element multiplication operation, and M is a mask matrix with the same shape as W. When the absolute value of the elements in W is less thanθ , M at the corresponding position is 0, and vice versa is 1. The importance of the weights is determined by calculating their gradient contribution during training, and the weights with smaller gradient contribution will be pruned after a threshold is set. The calculation formula can be expressed as$I(w) = \frac{|dw|}{\sum|dw|}$, where I(w) is the importance index of weight w, and dw is the gradient of the weight in the backpropagation process. The weights whose importance is below the threshold are selected for pruning. In addition, for the convolutional layer, also we directly prune the whole filter (Filter), i.e., if the output of a filter has less impact on the loss function, the whole filter and its corresponding weights are deleted. By model pruning, not only can the model size be effectively compressed, but also reduce the amount of computation, which is conducive to the efficient implementation of real-time target detection on embedded devices.

4.3. Adaptive environment sensing and control strategy

1) *Adaptive Algorithm Design for Complex Situations such as Light Changes and Occlusion*

Under varying lighting conditions, light invariant feature extraction algorithms such as LocalBinaryPatterns (LBP) or HistogramEqualization are used to enhance the image contrast and resist the effects of varying light intensity. For image preprocessing, histogram equalization can be expressed as: $I_{eq}(x,y) = T[h(I(x,y))]$. Where$I(x,y)$ represents the original image pixel values, h is the cumulative distribution function (CDF), and T is the inverse mapping function that maps the uniform distribution back to the original pixel value interval. When occlusion is encountered, deep learning based occlusion recovery algorithms such as Generative Adversarial Networks (GAN) or Recurrent Neural Networks (RNN) can be used to predict and fill in the missing information. The learning objective function of the occlusion recovery model can be expressed as follows: $\mathcal{L} = \mathcal{L}_{data}(G(X_{occ}), X_{gt}) + \lambda \mathcal{L}_{adv}(D(G(X_{occ})))$ where G is the generator for predicting the content of the occluded portion,X_{occ} is the occluded image,X_{gt} is the unoccluded ground truth image, D is the discriminator,$\mathcal{L}_{data}$ is the reconstruction loss,$\mathcal{L}_{adv}$ is the adversarial loss, andλ is the weight coefficient that balances the two parts of the loss.

2) *Dynamic target tracking and real-time feedback control mechanisms*

Dynamic target tracking is the process of locating, predicting and tracking a moving target in a real-time scene. In this process, algorithms such as Kalman filtering and particle filtering are used to predict and track the motion trajectory of the target in real time, so as to realize the precise positioning and real-time tracking of the target.

Kalman filtering is a linear optimal estimation algorithm that achieves optimal prediction of the state of a system by recursively estimating the state of the system in a linear dynamic system. In the prediction step: $\hat{x}_{k|k-1} = A\hat{x}_{k-1|k-1} + Bu_k$ $P_{k|k-1} = AP_{k-1|k-1}A^T + Q$. In the updating step: $K_k = P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1}$ $\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - H\hat{x}_{k|k-1})$, where $\hat{x}_{k|k-1}$ and $\hat{x}_{k|k}$ are the predicted and updated state estimates, respectively, $P_{k|k-1}$ and $P_{k|k}$ are the corresponding state covariance matrices, K_k is the Kalman gain, A is the state transfer matrix, B is the control input matrix, H is the observation matrix, Q and R are the covariance matrices of the process noise and the observation noise, respectively, andz_k is the current observation. In terms of the real-time feedback control mechanism, the target tracking results are fed back to the actuator (e.g., robot or unmanned vehicle) to adjust its attitude and motion strategies through methods such as PID controller or model predictive control (MPC). For example, the output of the PID controller can be expressed as: $u(k) = K_p[e(k)] + K_i \int_0^t e(\tau)d\tau + K_d \frac{de(k)}{dt}$ where u(k) is the control output, e(k) is the current position error, and , , and $K_pK_iK_d$ are the proportional, integral, and differential gains, respectively.

5. EXPERIMENTAL EVALUATION

5.1. Experimental environment and platform construction

In order to comprehensively test the sensor-based automatic target recognition and detection system, this experiment is conducted in both simulated and real environments. The experimental environments cover different lighting conditions (daytime, nighttime, shadows, glare interference, etc.), multiple occlusion levels, and complex dynamic scenes. The experimental platform is equipped with a high-performance GPU server to run deep learning models, and various types of sensors, including but not limited to industrial-grade color and infrared cameras, VelodyneHDL-64E LIDAR, ultrasonic arrays, and high-quality IMU devices.

5.2. Experimental program

This experiment is designed to test the target detection task in multiple scenarios for targets of different sizes, speeds and materials. In each scenario, the system performs synchronized data acquisition through integrated multi-sensors and processes the raw data using the proposed data preprocessing and feature extraction techniques. Then, the feature-fused data are fed into an improved target detection model based on FasterR-CNN for training and testing.

The experiments firstly verified the synchronization accuracy between the sensors and the effectiveness of data fusion. A series of benchmark tests are designed to record and compare the target detection accuracy and robustness of different sensors working individually with multi-sensor fusion, especially the improvement of the adaptive ability to complex environments. Various preprocessing techniques and feature extraction methods are applied to image sensors and point cloud data, respectively, to compare the effects of different algorithms on target detection performance. The experiments documented the improvement of data quality after denoising, correction, and smoothing processes, as well as the enhancement effect of different feature extraction methods on target detection accuracy and recall. For the FasterR-CNN framework and its improved version, the performance on different scale target detection tasks is compared by adjusting the network structure and training parameters. The experiments pay special attention to the improvement of small target detection rate before and after the introduction of feature pyramid network (FPN), as well as the effects of attention

mechanism and model pruning on the overall performance and real-time performance.

5.3. Experimental implementation

The implementation of the experiments strictly followed the following steps: first, data collection was carried out using an integrated sensor system under different environmental conditions, while ensuring strict synchronization of all sensor data. Next, the collected datasets were manually labeled, and these labeled data will be used for subsequent deep learning model training and validation. On the labeled dataset, the FasterR-CNN and its improved version of the model were trained, and the effectiveness of the various modules of the model, such as Region Proposition Network (RPN), Region of Interest Pooling (RoIPooling), and Feature Pyramid Network (FPN), was independently validated.

5.4. Experimental results

Table 1 demonstrates the target detection performance comparison between single-sensor work and multi-sensor data fusion. The fused system shows significant improvement in both precision and recall, and the overall performance is enhanced despite the decrease in processing speed brought about by the fusion.

Table 1. Comparison of single sensor and fused target detection performance

Sensor type	Work on one's own	Post-fusion
RGB Camera	Precision:85%	Precision:92%
	Recall:78%	Recall:88%
infrared camera	Precision:80%	Precision:87%
	Recall:72%	Recall:80%
LiDAR	F1Score:83%	F1Score:90%
Consolidation of data	FPS (single sensor): 20fps	FPS(Fusion):15fps

Table 2 shows the effect of different image preprocessing methods on sensor performance. Denoising and correction can improve the image quality and thus the accuracy of target detection.

Table 2. Effect of different preprocessing methods on image sensor performance

Pre-processing methods	Image accuracy after noise removal	Accuracy after image correction	Increased accuracy after smoothing
median filter	+3%	-	+1%
Gaussian filter	+2%	-	+2.5%
distortion correction	-	+5%	-
Projection Transformation Correction	-	+2%	-

Table 3 illustrates the impact of different feature extraction methods on target detection performance. The image sensor improves precision and recall through edge and corner detection, while the LiDAR point cloud improves F1 score through point cloud clustering and dimensionality reduction.

Table 3. Effect of feature extraction methods on target detection performance

Feature Extraction Methods	image sensor	LiDAR point cloud
Edge detection (Sobel/Canny)	Precision↑4%	-
Corner point detection (Harris/SIFT)	Recall↑3%	-
point cloud clustering	-	F1Score ↑5%
PCA downscaling	-	F1Score ↑2%

Table 4 compares the performance of FasterR-CNN and its improved versions on different scale target detection tasks. The introduction of FPN and attention mechanism improves the detection accuracy, while model pruning increases the processing speed while maintaining higher accuracy.

Table 4. Comparison of the performance of the improved version of the fasterr-CNN model

Model structure	Large Target Detection	Mid-target detection	Small Target Detection	Real-time performance (FPS)
Original FasterR-CNN	Precision: 90%	Precision: 85%	Precision: 70%	10fps
With the introduction of FPN	Precision: 92%	Precision: 88%	Precision: 75%	8fps
Incorporation of attention mechanism	Precision: 93%	Precision: 89%	-	7fps
After model pruning	Precision: 92%	Precision: 87%	Precision: 73%	12fps

Table 5 shows the target detection performance of RGB camera, IR camera and LiDAR under different environmental conditions. The RGB camera performs best during the daytime, while the IR camera has higher accuracy at night. The LiDAR performs stably under all conditions, but its performance is slightly degraded under strong light interference.

5.4. Discussion

In this study, the automatic target recognition and detection system based on improved machine vision control demonstrated excellent performance under various experimental conditions. Through the

comparative analysis of the data in Tables I to V, we can summarize the following points.

Table 5. Comparison of target detection performance under different environmental conditions

Environmental conditions	Daytime	Evening	Fig. a traumatic experience that haunts someone	Glare
RGB Camera Accuracy	90%	75%	85%	70%
Infrared Camera Accuracy	N/A	85%	N/A	N/A
LiDARF1 Score	90%	88%	85%	80%

Advantages of multi-sensor data fusion: As can be seen from Table I, compared with the performance of a single sensor working independently (such as RGB camera accuracy 85%, recall rate 78%; infrared camera accuracy 80%, recall rate 72%), the fused system significantly improves the accuracy and recall rate of target detection (reaching 92% and 88%, respectively). Although the processing speed has decreased (from 20fps to 15fps), the overall performance improvement proves the effectiveness of multimodal information integration.

Influence of preprocessing methods: Table II reveals the specific effects of different image preprocessing techniques on sensor performance. For example, after removing noise using median filtering or Gaussian filtering, the image accuracy is improved by 3% and 2%, respectively; distortion correction further improves the accuracy by 5%. These results show that appropriate preprocessing steps are crucial to enhance data quality in subsequent feature extraction and target detection processes.

Effect of feature extraction methods: Table III shows that edge detection (such as Sobel/Canny) and corner detection (such as Harris/SIFT) for image sensors can effectively improve the accuracy and recall of target detection; for LiDAR point cloud data, point cloud clustering and PCA dimensionality reduction technology can significantly improve the F1 score, indicating that choosing a suitable feature extraction strategy is of great significance for different types of sensor data.

Benefits of model structure improvement: According to Table IV, after introducing feature pyramid network (FPN) and attention mechanism on the basis of the original Faster R-CNN, the detection accuracy of large, medium and small targets has been improved to varying degrees. In particular, the application of FPN has significantly enhanced the recognition ability of small targets (from 70% to 75%). At the same time, the model pruning technology not only retains a high accuracy rate, but also greatly improves the real-time processing speed (from 10fps to 12fps), which is particularly beneficial for the actual deployment of embedded devices.

Consideration of environmental adaptability: Finally, Table V shows the target detection performance under different lighting conditions. RGB cameras perform best during the day, while infrared cameras are more suitable for night scenes. In contrast, LiDAR exhibits good all-weather stability, but its performance slightly decreases under strong light interference. This suggests that we need to consider the impact of environmental factors on the working status of various types of sensors when designing practical application systems, and take corresponding measures to optimize the overall performance.

In summary, through comparative analysis of existing methods, we found that the system shows great potential in target recognition and detection in complex environments. However, future research still needs to focus on issues such as expanding the scale of training sample sets and coping with extreme weather conditions to further improve the robustness and practicality of the system.

6. CONCLUSION

In this study, an automatic target recognition and detection system for sensors based on machine vision control was successfully designed and implemented. The system achieves high-precision detection of targets in complex environments by integrating multiple sensors, including industrial-grade color and infrared cameras, VelodyneHDL-64E lidar, ultrasonic arrays, and high-quality IMU devices. Experimental results show that multi-sensor data fusion significantly improves the accuracy and robustness of target detection, especially in complex environments and dynamic scenes. A series of benchmark tests are designed to validate the synchronization accuracy between sensors and the effectiveness of data fusion, as well as the performance enhancement under various preprocessing techniques and feature extraction methods. The experiments also demonstrate the excellent performance of FasterR-CNN and its improved version of the model on target detection tasks at different scales, especially with the introduction of the feature pyramid network (FPN) and the attention mechanism, which results in a significant increase in the detection rate and the overall performance of small targets. In addition, the model pruning technique successfully realizes the lightweighting of the model while maintaining high detection accuracy, which provides the possibility of real-time target detection for embedded devices.

Despite the remarkable results, there are still some limitations in this study. First, the relatively small size of the dataset in the experiment may not fully cover all possible complex scenarios and target types. Second, the performance of the system may be affected under certain extreme conditions,

such as extreme light variations or strong occlusions. In addition, due to the limitation of experimental conditions, the validation of the target tracking and real-time feedback control mechanism in certain dynamic scenes is not sufficient. Finally, the real-time performance of the model still has room for improvement in some cases, especially in terms of processing speed.

Future research work can be carried out in the following aspects: first, further expand the size of the dataset by adding more different types of targets and complex scenarios to improve the generalization ability of the system. Second, more robust preprocessing techniques and feature extraction methods are explored to improve the performance of the system under extreme conditions. In addition, target tracking and real-time feedback control mechanisms under dynamic scenes can be studied in depth to achieve a more intelligent and flexible system response.

Source of funding: *This research received no external funding.*

Author contributions: *research concept and design, W.Z.; Collection and/or assembly of data, W.Z., Y.X., Y.H.; Data analysis and interpretation, Y.X., Y.H.; Writing the article, W.Z., Y.X., Y.H.; Critical revision of the article, W.Z., Y.X., Y.H.; Final approval of the article, W.Z., Y.X., Y.H.*

Declaration of competing interest: *The author declares no conflict of interest.*

REFERENCES

1. Kechagias-Stamatis O, Aouf N. A new passive 3-D automatic target recognition architecture for aerial platforms. IEEE Transactions on Geoscience and Remote Sensing. 2019;57(1):406-15. https://doi.org/10.1109/TGRS.2018.2855065.
2. Kim SH, Choi HL. Convolutional neural network-based multi-target detection and recognition method for unmanned airborne surveillance systems. International Journal of Aeronautical and Space Sciences. 2019;20(4):1038-46. https://doi.org/10.1007/s42405-019-00182-5.
3. Gong YM, Ma ZY, Wang MJ, Deng XY, Jiang W. A new multi-sensor fusion target recognition method based on complementarity analysis and neutrosophic set. Symmetry-Basel. 2020;12(9). https://doi.org/10.3390/sym12091435.
4. So HY, Kim EM. Deep-learning-based automatic detection and classification of traffic signs using images collected by mobile mapping systems. Sensors and Materials. 2022;34(12):4801-12. https://doi.org/10.5194/isprs-archives-XLVIII-4-W9-2024-183-2024.
5. Dai J, Hao XH, Yan XP, Li Z. Adaptive false-target recognition for the proximity sensor based on joint-feature extraction and chaotic encryption. IEEE Sensors Journal. 2022;22(11):10828-40. https://doi.org/10.1109/JSEN.2022.3169746.
6. Teixeira E, Araujo B, Costa V, Mafra S, Figueiredo F. Literature review on ship localization, classification, and detection methods based on optical sensors and neural networks. Sensors. 2022;22(18). https://doi.org/10.3390/s22186879.
7. Blomerus N, Cilliers J, Nel W, Blasch E, de Villiers P. Feedback-assisted automatic target and clutter discrimination using a bayesian convolutional neural network for improved explainability in SAR applications. Remote Sensing. 2022;14(23). https://doi.org/10.3390/rs14236096.
8. Wang SJ, Jiang F, Zhang B, Ma R, Hao Q. Development of UAV-based target tracking and recognition systems. IEEE Transactions on Intelligent Transportation Systems. 2020;21(8):3409-22. https://doi.org/10.1109/TITS.2019.2927838.
9. Gao JC, Zhang XQ. An information recognition and time extraction method of tracking a flying target with a sky screen sensor based on wavelet modulus maxima theory. Mathematics. 2023;11(18). https://doi.org/10.3390/math11183936.
10. Li X, Zheng H. Target detection algorithm for dance moving images based on sensor and motion capture data. Microprocessors and Microsystems. 2021;81. https://doi.org/10.1016/j.micpro.2020.103743.
11. Bin KC, Lin J, Tong XQ, Zhang XP, Wang JQ, Luo SH. Moving target recognition with seismic sensing: A review. Measurement. 2021;181. https://doi.org/10.1016/j.measurement.2021.109584.
12. Liang B, Wang X, Zhao WH, Wang XB. High-precision carton detection based on adaptive image augmentation for unmanned cargo handling tasks. Sensors. 2024;24(1). https://doi.org/10.3390/s24010012.
13. Zheng XT. Laser radar-based intelligent vehicle target recognition and detection system using image detection technology. Journal of Electronic Imaging. 2023;32(1). https://doi.org/10.1117/1.JEI.32.1.011203.
14. Li K, Zhang YS, Zhang ZC, Yu Y. An automatic recognition and positioning method for point source targets on satellite images. Isprs International Journal of Geo-Information. 2018;7(11). https://doi.org/10.3390/ijgi7110434.
15. He XC, Ji WL, Xing SJ, Feng ZX, Li HY, Lu SS, et al. Emerging trends in sensors based on molecular imprinting technology: Harnessing smartphones for portable detection and recognition. Talanta. 2024; 268. https://doi.org/10.1016/j.talanta.2023.125283.
16. Li B, Zhou SS, Cheng LF, Zhu RB, Hu T, Anjum A, et al. A cascade learning approach for automated detection of locomotive speed sensor using imbalanced data in ITS. IEEE Access. 2019;7:90851-62. https://doi.org/10.1109/ACCESS.2019.2928224.
17. Samadiani N, Huang GY, Cai BR, Luo W, Chi CH, Xiang Y, He J. A review on automatic facial expression recognition systems assisted by multimodal sensor data. Sensors. 2019;19(8). https://doi.org/10.3390/s19081863.
18. Choudhary P, Goel N, Saini M. A survey on seismic sensor based target detection, localization, identification, and activity recognition. Acm Computing Surveys. 2023;55(11). https://doi.org/10.1145/3568671.
19. Seng KP, Ang LM, Schmidtke LM, Rogiers SY. Computer vision and machine learning for viticulture technology. IEEE Access. 2018;6:67494-510. https://doi.org/10.1109/ACCESS.2018.2875862.
20. Li DL, Wang Q, Li X, Niu ML, Wang H, Liu CH. Recent advances of machine vision technology in fish classification. Ices Journal of Marine Science.

2022;79(2):263-84. https://doi.org/10.1093/icesjms/fsab264.
21. Hsia SC, Wang SH, Wei CM, Chang CY. Intelligent object tracking with an automatic image zoom algorithm for a camera sensing surveillance system. Sensors. 2022;22(22). https://doi.org/10.3390/s22228791.
22. Fu Q, Wang ST, Wang J, Liu SN, Sun YB. A Lightweight eagle-eye-based vision system for target detection and Recognition. IEEE Sensors Journal. 2021;21(22):26140-8. https://doi.org/10.1109/JSEN.2021.3120922.
23. Zhang WY, Fu XH, Li W. The intelligent vehicle target recognition algorithm based on target infrared features combined with lidar. Computer Communications. 2020;155:158-65. https://doi.org/10.1016/j.comcom.2020.03.013.
24. Zabit U, Shaheen K, Naveed M, Bernal OD, Bosch T. Automatic detection of multi-modality in self-mixing interferometer. IEEE Sensors Journal. 2018;18(22):9195-202. https://doi.org/10.1109/JSEN.2018.2869771.
25. Wang HL, He S, Yu JS, Wang LY, Liu T. Research and implementation of vehicle target detection and information recognition technology based on NI myRIO. Sensors. 2020;20(6). https://doi.org/10.3390/s20061765.
26. Hou B, Zhang CP, Yang SB. Computer vision tool-setting system of numerical control machine tool. Sensors. 2020;20(18). https://doi.org/10.3390/s20185302.
27. Xing KS, Wang N, Wang W. A ground moving target detection method for seismic and sound sensor based on evolutionary neural networks. Applied Sciences-Basel. 2022;12(18). https://doi.org/10.1109/LGRS.2021.3063767.
28. Wei PC, Wang B. Multi-sensor detection and control network technology based on parallel computing model in robot target detection and recognition. Computer Communications. 2020;159:215-21.
29. Liu JM, Chen H, Wang Y. Multi-source remote sensing image fusion for ship target detection and recognition. Remote Sensing. 2021;13(23). https://doi.org/10.3390/rs13234852.
30. Kotwaliwale N, Singh K, Chakrabarty SK, Joshi MA, Kalne A, Tiwari M, et al. Machine vision for characterisation of some phenomic features of plant parts in distinguishing varieties-a review. International Journal of Bio-Inspired Computation. 2019;14(4):201-12. https://doi.org/10.1504/IJBIC.2019.103960.
31. Long T, Liang ZN, Liu QH. Advanced technology of high-resolution radar: target detection, tracking, imaging, and recognition. Science China-Information Sciences. 2019;62(4). https://doi.org/10.1007/s11432-018-9811-0.
32. Baili N, Moalla M, Frigui H, Karem AD. Multistage approach for automatic target detection and recognition in infrared imagery using deep learning. Journal of Applied Remote Sensing. 2022;16(4). https://doi.org/10.1117/1.JRS.16.048505.
33. Geng Z, Xu Y, Wang BN, Yu X, Zhu DY, Zhang G. Target recognition in SAR images by deep learning with training data augmentation. Sensors. 2023;23(2). https://doi.org/10.3390/s23020941.
34. Radcliffe J, Cox J, Bulanon DM. Machine vision for orchard navigation. Computers in Industry. 2018;98:165-71. https://doi.org/10.1016/j.compind.2018.03.008.
35. Gebauer J, Sofer P, Jurek M, Wagnerová R, Czebe J. Machine vision-based fatigue crack propagation system. Sensors. 2022;22(18). https://doi.org/10.3390/s22186852.
36. Wu L, Gong FX, Yang X, Xu L, Chen SY, Zhang Y, et al. The YOLO-based multi-pulse lidar (YMPL) for target detection in hazy weather. Optics and Lasers in Engineering. 2024;177. https://doi.org/10.3390/electronics13010220.
37. Cheng C, Gao M, Cheng XD, Fang D, Chen YC. Research on fast target recognition method based on spectrum detection in battlefield. Spectroscopy and Spectral Analysis. 2018;38(1):161-5. https://doi.org/10.3964/j.issn.1000-0593(2018)01-0161-05.
38. Pan QC, Zhang HH. Key algorithms of video target detection and recognition in intelligent transportation systems. International Journal of Pattern Recognition and Artificial Intelligence. 2020;34(9). https://doi.org/10.3390/app13063995.
39. Song YF, Dong GG. Federated target recognition for multiradar sensor data security. IEEE Transactions on Geoscience and Remote Sensing. 2023;61.
40. Chen WS, Chen XL, Liu J, Wang QB, Lu XF, Huang YF. Detection and recognition of UA targets with multiple sensors. Aeronautical Journal. 2023;127 (1308):167-92. https://doi.org/10.1017/aer.2022.50.

Wei ZHOU was born in Zhengzhou, Henan Province, China in 1975. She obtained a Bachelor of Science degree in Computer Science from Shaanxi Normal University in 2000, and obtained a Master's degree in Light Industry Information Technology and Engineering from Jiangnan University in 2009.

Since 2000, she has been working at Wuxi Institute of Technology and currently serves as an associate professor in the Department of Software Technology. She started working in wireless sensor networks and data encryption research, completed multiple horizontal projects for enterprises. She has published two books, multiple articles, and multiple technical patents. Among them, one article SCI articles, two core Chinese articles, and one invention patent.

Since 2018, she has led students to participate in professional competitions and won two first prizes in vocational education in Jiangsu Province and one second prize nationwide.
e-mail: zhou-wei163@hotmail.com

Ye XU received his PhD in signal and information processing from Communication University of China, Beijing, China, in 2017.

He is currently a faculty member of the school of IoT technology, Wuxi Institute of Technology. His research interests include computer vision, pattern recognition, and deep learning.
e-mail: Ye_Xu20@outlook.com

Yue HUANG received the Ph.D. degree in Light Industry Information and Engineering from the School of Internet of Things Engineering, Jiangnan University in 2012. From 2012 to 2019, she was a Lecturer with the School of Internet of Things Technology, Wuxi Institute of Technology.

Since 2019 has been an Associate Professor with the same institute.

She is the author of two books, more than 10 articles, more than 5 patents and 2 software copyrights. She has led a sub-project of national-level teaching resource library and a provincial-level scientific research project in Jiangsu Province.
e-mail: Yue_Huang20@outlook.com