Application of Visual Perception and Recognition Technology of Intelligent Inspection Robots in Guanyinyan Hydropower Station

Youliang He*[1], Zhiguo Ma, Chunlei Li, Zefa Xin, Weimin Ouyang
Datang Guanyinyan Hydropower Development Co., Ltd, China

ABSTRACT

In order to ensure the safe operation of hydropower stations, a device state extraction and recognition method based on visual perception of inspection robots is proposed. The experimental results show that the proposed device state extraction and recognition method based on visual perception of inspection robots has the lowest recognition accuracy for buttons, switches, text, and instrument readings, which are 97.2%, 96.8%, 97.7%, and 94.7%, respectively. The average accuracy is 97.9%, 97.4%, 98.7%, and 95.3%, respectively, which are higher than other algorithms. In addition, the area under the working characteristic curve of the proposed method is 0.97, which is higher than other algorithms, and its accuracy recall curve completely envelops other algorithms. The above results indicate that the proposed equipment state extraction and recognition method based on visual perception of inspection robots can accurately identify the status of different equipment, providing strong support for the safe operation of hydropower stations.

Keywords: Inspection robots; Visual perception; State extraction; Differentiatable binary network; Convolutional Recurrent Neural Network

1. INTRODUCTION

Hydropower stations are of great significance in solving energy shortages, improving the ecological environment, promoting regional economic coordination and sustainable development. They not only provide a large amount of clean and inexpensive energy, but also play a role in ensuring agricultural development and preventing flood disasters. Hydroelectric power stations are comprehensive engineering facilities that convert water energy into electrical energy, serving as frequency regulation, peak shaving, phase regulation, backup and other tasks in the power system. The main facilities of a hydropower station include reservoirs formed by water retaining and discharge structures, as well as the water diversion system of the hydropower station, power plants, mechanical and electrical equipment, etc. Once various electrical equipment malfunctions, it will seriously threaten the operational safety of the hydropower station [1-2]. Therefore, in order to ensure the operation and maintenance safety of hydropower stations, it is necessary to conduct regular inspections. However, due to the low efficiency and significant safety hazards of manual inspections, it is difficult to detect equipment problems in a timely manner. Therefore, in order to achieve monitoring of the status of hydropower station equipment, a method for extracting and recognizing equipment status based on visual perception technology of inspection robots has been proposed. This method innovatively constructs state extraction algorithms for different types of devices, and studies the use of Hue, Saturation, Value (HSV) color space and region bounding ellipses for state extraction of control devices; For text, research has combined Differential Binary Network (DBNet) and Convolutional Recurrent Neural Network (CRNN) for recognition; For instrument readings, research utilizes threshold segmentation and line detection for recognition.

2. ALGORITHM FOR EXTRACTING THE STATUS OF HYDROPOWER STATION EQUIPMENT BASED ON ROBOT VISION

2.1 Algorithm for identifying the status of buttons and switches in hydropower station equipment

When using inspection robots to monitor the electrical equipment of hydropower stations, due to the wide variety of equipment types, it is necessary for robots to accurately distinguish the status of all types of equipment. In view of this, research has designed state information extraction algorithms for different types of devices. For device controllers, they

[1] *johnevans@163.com; jack_ou1994@163.com;jianguoquan@163.com; yangmingjie0305@163.com; suyouershang@163.com

The International Conference Optoelectronic Information and Optical Engineering (OIOE2024), edited by Yang Yue, Lu Leng, Proc. of SPIE Vol. 13513, 135133A · © 2025 SPIE · 0277-786X · © The Authors.
 · doi: 10.1117/12.3048496

generally include two types: buttons and switches. For buttons, their status is generally reflected through color brightness. Therefore, in order to accurately reflect the status of device buttons, a button status recognition algorithm based on HSV color space was proposed. Compared to the Red Green Blue (RGB) color space, the HVS color space can represent colors through shooting, saturation, and brightness. The HVS color space model is shown in Figure 1.

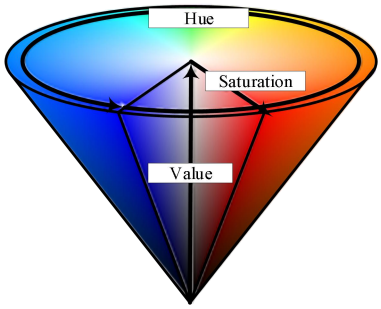

Figure 1 HVS color space model

As shown in Figure 1, in the HVS color space model, the color tone starts to take values counterclockwise, with a range of 0-360 °. Taking red as an example, its corresponding angle is 150 °, while green corresponds to an angle of 30 °. The saturation starts from the center of the model, which has a saturation of 0% and appears white; The further outward, the higher the saturation value and the darker the color. The brightness starts from the lower end, where the brightness is 0% and appears black. The higher the brightness, the brighter the color, which tends towards white [4-5]. However, as the image color model used by the inspection robot is an RGB color space model, it needs to be converted. The calculation formula for converting the RGB color space model to the HSV color space model is shown in equations (1) to (3).

$$V = \frac{\max(R,G,B)}{255} \tag{1}$$

In equation (1), V represents brightness; $\max(R,G,B)$ Represents the maximum value of a pixel in RGB space.

$$S = \begin{cases} V - \dfrac{\min(R,G,B)}{255}, & V \neq 0 \\ \dfrac{V}{0}, & V = 0 \end{cases} \tag{2}$$

In equation (2), S represents saturation; $\min(R,G,B)$ Represents the minimum value of a pixel in RGB space.

$$H = \begin{cases} \dfrac{(G-B)}{\dfrac{255}{S}} \times 60°, & V = R/255 \\ 120° + \dfrac{(B-R)}{\dfrac{255}{S}} \times 60°, & V = G/255 \\ 240° + \dfrac{(R-G)}{\dfrac{255}{S}} \times 60°, & V = B/255 \end{cases} \tag{3}$$

In equation (3), H represents the color tone G And B respectively represent red, green, and blue. After converting the RGB color space model of the original image to the HVS color space model, the state of the device buttons can be

determined based on the distribution of brightness values in the image. For switches, as they are generally operated in a rotating manner, their status can be determined by the angle of the switch. When extracting the status information of a switch, the first step is to determine the contour area of the switch. Therefore, in order to accurately fit the contour area of the switch, the method of region bounding ellipse was adopted in the study. At the same time, the angle between the major axis of the ellipse and the horizontal line is used to reflect the state of the switch. Based on the contour point information of the switch area, it is possible to calculate the major and minor axes and centers of the minimum area's outer ellipse. The relationship between the coefficients of the minimum region bounding ellipse and the second-order row moment, column moment, and second-order mixed matrix is shown in equation (4).

$$\begin{vmatrix} d & e \\ e & f \end{vmatrix} = \frac{1}{3\left(\mu_x \mu_y - \mu_{xy}^2\right)} \begin{vmatrix} \mu_x & -\mu_{xy} \\ -\mu_{xy} & \mu_y \end{vmatrix} \tag{4}$$

In equation (4), d , f Both e and represent the coefficients of the minimum region's bounding ellipse; μ_x represents the second-order row moment of the pixel region; μ_y Represent second-order column moments; μ_{xy} Represents a second-order mixed matrix. By using equation (6), the coefficient matrix of the minimum region's bounding ellipse can be calculated to obtain the elliptic equation, and ultimately determine the major and minor axes and center of the minimum region's bounding ellipse. When extracting the status information of the switch, the long axis direction of the minimum area outer ellipse can be used as the rotation direction of the switch, where the positive direction is counterclockwise and the negative direction is clockwise. The state of the switch can be determined based on the angle between the long axis and the horizontal line.

2.2 Recognition algorithm for text and reading of hydropower station equipment

In hydropower station equipment, in addition to various switch buttons, there are also various instruments and meters that often reflect the equipment status through text or readings. Therefore, in addition to extracting the status of control devices, inspection robots also need to extract the text or readings of instruments and meters. For text extraction, it involves two steps: text detection and recognition. For text detection based on robot vision, existing methods are limited to complex post-processing algorithms and the scale robustness of their segmentation models, resulting in low detection accuracy and long processing time. Therefore, in order to achieve accurate and fast text detection, research is being conducted using DBNet as a text detection algorithm. The DBNet algorithm integrates the binarization process into the segmentation network and introduces an effective adaptive scale fusion module to improve scale robustness by adaptively fusing features of different scales. The structure of DBNet is shown in Figure 2.

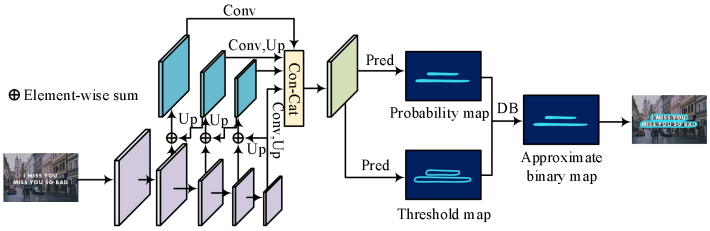

Figure 2 Structure of DBNet

As shown in Figure 1, DBNet first performs multiple feature extraction and upsampling on the input image, and then fuses the obtained feature maps at all levels to obtain the final feature map. Then, based on the fused feature map, predict the probability map and threshold map, and calculate the approximate binary map of the two. Finally, the approximate binary image is fused with the input image to obtain the text detection result [6]. The approximate binarization calculation formula for probability graphs is shown in equation (5).

$$\hat{B}_{i,j} = \frac{1}{1 + e^{-k\left(P_{i,j} - T_{i,j}\right)}} \tag{5}$$

In equation (5), $\hat{B}_{i,j}$ represents an approximate binary plot; (i,j) Represent the pixel coordinates in the probability plot; k Gain factor of surface gradient; $P_{i,j}$ Represent a probability graph; $T_{i,j}$ Represents a threshold graph after network optimization. For text recognition, the algorithm used in the research is CRNN, which includes three parts: convolutional layer, loop layer, and transcription layer. The convolutional layer is responsible for feature extraction of the input image, and it should be noted that the input image at this time is only the text area in the original image; The recurrent layer is responsible for learning feature vectors through a bidirectional recurrent neural network and obtaining predicted label distributions; The transcription layer is responsible for converting the predicted label distribution into label sequences [7-8].

In addition to recognizing various types of text, inspection robots also need to recognize instrument readings. Therefore, in order to achieve accurate recognition of instrument readings, the study used threshold segmentation to segment the pointer of the instrument and extract its angle. Before performing threshold segmentation, it is necessary to first grayscale the image. Next, threshold segmentation can be performed. The threshold segmentation method used in the study is the maximum inter class variance method, which divides the image into background and target parts based on its grayscale characteristics, and obtains the segmentation threshold by maximizing the inter class variance [9-10]. The formula for calculating the grayscale probability of an image is shown in equation (6).

$$P_i = \frac{n_i}{MN} \tag{6}$$

In equation (6), P_i represents the i probability that the grayscale level is; The n_i grayscale level is the i number of pixels; The $M \times N$ total number of pixels MN in an image of size. The formula for calculating inter class variance and partition threshold is shown in equation (7).

$$\begin{cases} \gamma = P_B(m)\left(V_B(m) - V_m\right)^2 + P_O(m)\left(V_O(m) - V_m\right)^2 = P_B(m)P_O(m)\left(V_B(m) - V_O(m)\right)^2 \\ \gamma(T) = \max\limits_{0 \le m \le L-1} \gamma(m) \end{cases} \tag{7}$$

In equation (13), γ represents the inter class variance; V_m Represents the average grayscale value of an image; T Indicates the partition threshold. After threshold partitioning, an image containing the contour of the instrument pointer can be obtained, and then Hough line detection is used to extract the features of the pointer contour. Assuming the existence of a straight line in a Cartesian coordinate system, the parameter equation of the straight line is shown in equation (8).

$$x\cos\theta + y\sin\theta = \rho \tag{8}$$

In equation (8), (x,y) represents the coordinates of points on a straight line; θ Represents the angle between the perpendicular line passing through the origin and the x-axis; ρ Represents the distance between a straight line and the origin. Map any point θ in a straight line to a ρ - coordinate system, where the point is mapped as a line, and the coordinates of each line relative to a point are the parameters of the line equation. By using the above method, the contour of the pointer can be approximated as a straight line, and then combined with the range of the instrument, the instrument reading can be extracted.

3. EQUIPMENT STATUS RECOGNITION RESULTS BASED ON ROBOT VISION

To verify the performance of the proposed robot vision based device state extraction algorithm, it was tested and compared with improved YOLOv5, path aggregation network optimization YOLOv3 (YOLOv3 for Path Aggregation Network Optimization, YOLOv3 PANet), YOLOv3 CRNN, and Dual path heterogeneous convolutional neural network

(DPHCNN). The dataset used in the experiment is a self built dataset, collected from Guanyinyan Hydropower Station. This dataset contains a total of 2098 images of various devices, with a ratio of 8:1:1 for the training, validation, and testing sets. The CPU used in the experiment is Intel Core i5 6800K 32GB, GPU is NVIDIA GeForce GTX 1080, and operating system is Ubuntu 18.04. The recognition results of each algorithm for the controller state are shown in Figure 3.

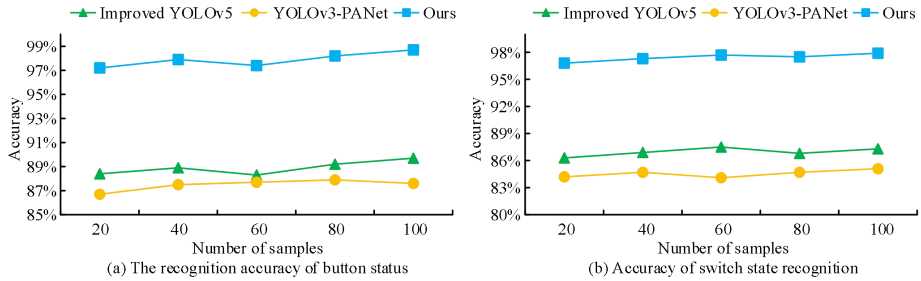

Figure 3 The recognition results of controller states by various algorithms

As shown in Figure 3 (a), the improved YOLOv5, YOLOv3 PANet, and the proposed method have the lowest recognition accuracy for button states, which are 88.4%, 86.7%, and 97.2%, respectively, with an average accuracy of 88.9%, 87.5%, and 97.9%, respectively. From Figure 3 (b), it can be seen that the improved YOLOv5, YOLOv3 PANet, and the proposed method have the lowest recognition accuracy for switch states, which are 86.3%, 84.1%, and 96.8%, respectively, with an average accuracy of 87.0%, 8%, 4.6%, and 97.4%, respectively. The above results indicate that for controller devices, the proposed method can accurately extract and recognize their states. The recognition results of different algorithms for text or reading are shown in Figure 4.

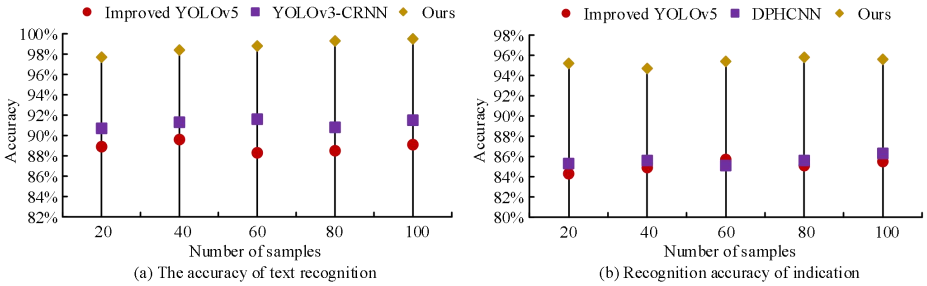

Figure 4 The recognition results of different algorithms for text or indication

From Figure 4 (a), it can be seen that the improved YOLOv5, YOLOv3 CRNN, and the proposed method have the lowest recognition accuracy for text at 88.3%, 90.7%, and 97.7%, respectively, with an average accuracy of 88.9%, 91.2%, and 98.7%, respectively. As shown in Figure 4 (b), the improved YOLOv5, DPHCNN, and the proposed method have the lowest recognition accuracy for instrument readings of 84.3%, 85.1%, and 94.7%, respectively, with average accuracy of 85.1%, 85.6%, and 95.3%. The above results indicate that the proposed method can accurately recognize text and instrument readings. The ROC and P-R curves of each algorithm are shown in Figure 5.

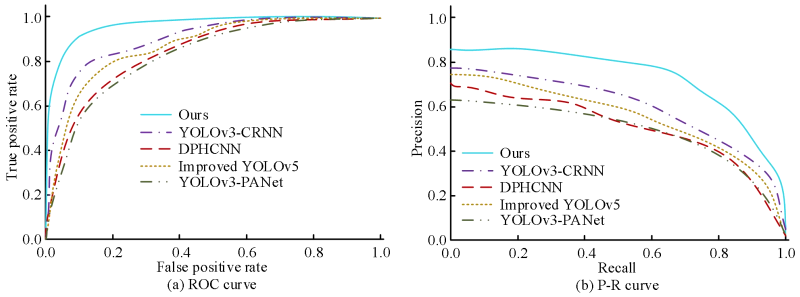

Figure 5 ROC curves and P-R curves of each algorithm

From Figure 5 (a), it can be seen that the ROC curves of the improved YOLOv5, YOLOv3 PANet, YOLOv3 CRNN, and DPHCNN have an area under the curve of 0.87, 0.84, 0.90, and 0.85, respectively. However, the ROC curves of the

proposed method completely envelop the ROC curves of other algorithms, with an area under the curve of 0.97. It can be seen that the algorithm proposed by the acting team has better state recognition performance. As shown in Figure 5 (b), compared to other algorithms, the P-R curve of the proposed algorithm is closer (1.0,1.0), indicating better performance of the proposed algorithm. The above results indicate that the overall performance of the proposed method is excellent.

4. CONCLUSION

Due to the presence of a large number of electrical equipment, hydropower stations require regular inspections to ensure safe operation. In view of this, the study proposes a method for extracting the status of hydropower station equipment based on visual perception of inspection robots. The experimental results show that the proposed method has the lowest recognition accuracy of 97.2% for button states, with an average accuracy of 97.9%; The lowest recognition accuracy for switch status is 96.8%, with an average accuracy of 97.4%; The lowest recognition accuracy for text is 97.7%, with an average accuracy of 98.7%; The recognition accuracy of instrument readings is the lowest at 94.7%, with an average accuracy of 95.3%, both higher than other algorithms. The above results indicate that the proposed method for extracting the status of hydropower station equipment can accurately identify the status of different equipment. However, the recognition of instrument readings requires the manual combination of range and pointer angle, and the algorithm has significant limitations. Therefore, in the future, machine vision will be used to simultaneously obtain readings from both range and pointer aspects.

5. REFERENCE

[1] Ji H, Cui X, Ren W, Liu L, & Wang W. Visual inspection for transformer insulation defects by a patrol robot fish based on deep learning. IET Science Measurement Technology, 2021, 15(7): 606-618.

[2] Jafari-Tabrizi A, Gruber D P, Gams A. Exploiting image quality measure for automatic trajectory generation in robot-aided visual quality inspection. The International Journal of Advanced Manufacturing Technology, 2024, 132(9): 4885-4901.

[3] Winston S J, Manivannan P V. Visual servoing based self-calibration of robotic inspection system using rigid body transformation parameters. International Journal of Mechatronics and Automation, 2022, 9(4): 192-209.

[4] Al a R M, Al j I A, Al R H T H S. Adaptive HDR Image Blind Watermarking Approach Based on Redundant Discrete Wavelet Transform. International journal of interactive mobile technologies, 2023, 17(10): 136-154.

[5] Khallaf F, El-Shafai W, El-Rabaie E S M, Soliman N F, & El-Samie F E A. A novel hybrid cryptosystem based on DQFrFT watermarking and 3D-CLM encryption for healthcare services. Frontiers of Information Technology & Electronic Engineering, 2023, 24(7): 1045-1061.

[6] Zhou X, Tang C, Huang P, Tian S, Mercaldo F, & Santone A. ASI-DBNet: An Adaptive Sparse Interactive ResNet-Vision Transformer Dual-Branch Network for the Grading of Brain Cancer Histopathological Images. Interdisciplinary Sciences: Computational Life Sciences, 2022, 15(1): 15-31.

[7] Wei Y, Gong Z, Yang S, Ye K, & Wen Y. EdgeCRNN: an edge-computing oriented model of acoustic feature enhancement for keyword spotting. Journal of Ambient Intelligence and Humanized Computing, 2022, 13(3): 1525-1535.

[8] Asatani N, Lu H, Kamiya T, Mabu S, & Kido S. Classification of Respiratory Sounds by scSE-CRNN from Triple Types of Respiratory Sound Images. Medical Imaging and Information Sciences, 2021, 38(4): 152-159.

[9] Zhou T, Wang Y, Zhang L, Chen B, & Yu X. Underwater Multitarget Tracking Method Based on Threshold Segmentation. IEEE Journal of Oceanic Engineering: A Journal Devoted to the Application of Electrical and Electronics Engineering to the Oceanic Environment, 2023, 48(4): 1255-1269.

[10] Luo Y, Sa J, Song Y, Jiang H, Zhang C, & Zhang Z. Texture classification combining improved local binary pattern and threshold segmentation. Multimedia tools and applications, 2023, 82(17): 25899-25916.