MM Science Journal | www.mmscience.eu
ISSN 1803-1269 (Print) | ISSN 1805-0476 (On-line)
Special Issue | TEAM 2024
Transdisciplinary and Emerging Approaches to Multidisciplinary Science
11.9.2024 – 13.9.2024, Ostrava, Czech Republic
DOI: 10.17973/MMSJ.2024_10_2024061

TEAM2024-00002

AUTOMATIC PANEL QUALITY MONITORING SYSTEM USING VISION-BASED ABNORMAL OBJECT DETECTION

THAI-VIET DANG[1]*, NGOC-TAM BUI[1]

[1]Department of Mechatronic, School of Mechanical Engineering, Hanoi University of Science and Technology, Hanoi 10000, Vietnam

*Corresponding author; e-mail: viet.dangthai@hust.edu.vn

Abstract

The rate of "Not Good" products due to the presence of abnormal objects during the quality control inspection of mobile phone's panels is as high as 38 percent without supervision and 19 percent with naked eye's inspection. To reduce the loss of small abnormal objects and improve the recognition rate of object detection of automatic panel quality monitoring, the paper proposes abnormal object recognition based on a template matching and subtract background technique. Firstly, the object recognition network based on convolutional neural networks (CNNs) is introduced initially. Then, the Canny Edge Detection method improves the image input quality. Next, template matching is utilized to discover things that depart from a standard pattern or a predetermined model. Models for background reduction will be implemented instantly into smartphone panel quality checking systems. Continually, a vision-based quality control (QC) strategy will be developed. Lastly, experimental results from a practical vision-based automatic anomalous object identification system demonstrate the viability and enhancement of the suggested strategy.

Keywords:

Abnormal object detection, background subtraction, convolutional neural networks (CNNs), template matching, vision-based method

1 INTRODUCTION

Vision-based techniques (VBMs) have been implemented in a variety of applications, including intelligent monitoring systems [Dang 2023a], security [Dang 2022], quality assurance, vision robots and intelligent transportation [Dang 2023b]. VBM aims to identify the region of interest quickly and precisely in each frame of a video stream. Numerous tracking obstacles (occlusion, lighting change, the dynamic motion and zoom change of the item, etc.) limit the development of a practical system for detecting small abnormal objects [Dang 2023c].

The process of template matching is employed to detect patterns or objects in an image that match a specified template. Numerous applications, including object categorization, object identification, motion estimation, 3D reconstruction, and anomalous object detection, make extensive use of this technology [Youngmo 2021]. To define template matching, the template image is compared to live images. Similarity metrics are computed as an integral component of the comparison process, including cross-correlation and mean squared error. A region of the target image with the greatest similarity value signifies the detection of the target object. Conventional template matching provides a map of pixel's similarity value. Then, the method presents susceptible matching situations, including shifting illumination, significant non rigid object deformations, and occlusion [Oron 2017]. Template matching strategies include scale invariant template matching, rotation invariant template matching, and multi template matching. The purpose of these versions is to enhance the illumination, rotation, and scaling robustness. Dou et a. proposed a robust image matching method based on SIFT (Scale invariant feature transform) to address large geometrical differences in acquired images [Dou 2018]. So, the method retrieved SIFT features from the candidate region and template to match them. Yet, the altered illumination conditions in the workplace reduced the accuracy of the vision approach. Subsequently, the iSIFT (illumination and Scale Invariant Feature Transform) descriptor is presented in [Tang 2019], which improves the precision of image feature matching in situations with variable light levels. In addition, the image with blurred edges received more feature points for the descriptor than the image with crisp edges. Bukala et al. provided a suitable denoising strategy for object recognition with HOG (Histogram of oriented gradient) features to mitigate the negative impacts of noise and distortion [Bukala 2020]. Thus, the performance of HOG features for detecting patterns in the real world was improved. Computer vision, including object identification, categorization, and object determination, has recently achieved a significant success in industrial applications [Nguyen 2023]. CNNs require huge amounts of training data and samples of obtained practical images that are less affected by object detection across live images to be successful at such tasks. Newer approaches have been effectively applied to the feature space formed by the convolutional layers of a CNN, leading to improved performance [Dang 2023d]. CNNs have a significantly larger capacity for data categorization than manually constructed features. To overcome the aforementioned issues, this study employs an automatic abnormal item identification system based on a template matching algorithm and a background subtraction technique. When objects are absent from the background, background subtraction is applied. Subtraction of the background is extremely sensitive to background noise and the fidelity of the background model. The Candy filter further refines the foreground mask and enhances the efficacy of the anomalous item identification system by employing the Candy edge finding method to surmount these obstacles. The automatic vision-based quality assurance system is successfully designed to solve the high NG with no management and naked eye's inspection as 38 and 19 percent, respectively. Therefore, the NG rate must be reduced to less than 3 percent.

The paper consists of the following sections: The Research Method section describes CNN's model, which utilizes template matching and an algorithm to detect anomalies. Next, Experiment Results describe the actual automatic panel quality monitoring system. The mechanical system's design incorporates testing procedures and an experiment module. Then, the results of the experiment indicate that the quality of the visual system increases the rate of anomaly detection. Finally, the Conclusion presents the results acquired to demonstrate the validity and improvement of the method outlined in the research.

2 RESEARCH METHOD

2.1 Object detection of quality monitoring

The topic of this study is the detection of small abnormal items, which enhances the quality monitoring system for smartphone screens. This study utilized Single Shot Multi Box Detector (SSD) [Liu 2015] with a training dataset derived from Microsoft Common Object in Context (MS COCO) dataset which contains over 2.5 million labeled instances in 328,000 images [Lin 2014]. Moreover, the authors collected five hundred photographs of abnormal objects were collected from the actual production process, and two hundred additional images were selected from other production units in the same field. SSD is also a CNN framework using 300 x 300 or 500 x 500-pixel images for the input and output. Then, to define the classification scores for each class of the classifiers. The hardware requires the following GPU (Graphics Processing Unit): 58 Fps (Frame per second) for the SSD300 and 23 Fps for the SSD500.

2.2 Object classification

This study utilized Residual Network (ResNet) to classify items [Mahaur 2023]. To achieve this advantage, skip connections are added for bypassing many convolutional layers. Though, the identity mappings will create residual blocks to be network shortcuts. Furthermore, to increase the training efficiency, MSR (Machine Specific Register) initialization [Lin 2022] will be combined with BN (Batch Normalization) [He 2015] to reduce vanishing gradient deterioration. Hence, learning extremely deep networks are much simpler without residual connections. Figure 1 depicts the ResNet architecture which includes residual blocks shown as (1):

$$x_{l+1} = x_l + F(x_l, W_l) \tag{1}$$

where x_l and x_{l+1} are the l^{th} residual block's input and output; F presents the residual's mapping.

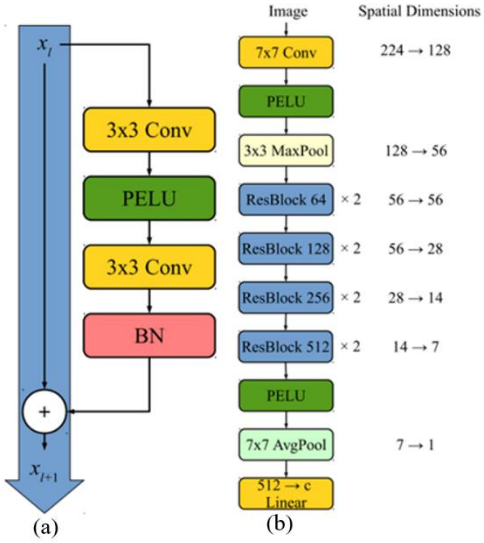


Fig. 1: The architecture of residual block and network [11] with (a): Residual block and (b): Residual network

As depicted in Figure 1a, the authors chose for a residual mapping, where an adaptive activation function PELU (Parametric Exponential Linear Unit) demonstrates superior performance on a variety of image classification tasks. Figure 1b depicts the 18-layer network with approximately 11.7 million learned parameters. It accepts a 224 x 224 color image input to generate a c-dimensional vector performing class conditional probabilities. The initial convolutional layer is comprised of 64: 7 x 7 filters convolved with a 2 x 2 stride and 3 x 3 zero padding. Then, this convolutional layer is denoted as Conv (6 x 4, 7 x 7, 2 x 2, 3 x 3). On the first layer's feature maps, PELU is constructed by a max pooling layer MaxPool (3 x 3, 2 x 2, 1 x 1). Then, in each convolutional layer, the input will travel through four residual blocks with 64, 128, 256, and 512 filters, respectively. The network then executes a final PELU activation of global average pooling AvgPool (7 x 7, 1 x 1). Linear layer 512 reaches to c, where c is the number of classes. Finally, a network increases the number of residual filter blocks, while the spatial dimension input is cut in half. This feature enables the network to gradually incorporate spatial data into the feature maps and retrieve higher level data. For example, the input spatial dimension of ResBlock 64 and ResBlock 128 is lowered from 56 x 56 pixels to 28 x 28 pixels, respectively. There is no identity skip connection because of the different dimensions of input and output. Therefore, the authors employ a convolutional layer with 1 x 1 filters and a stride of 2 x 2 to become a skip connection for the first ResBlock 128.

2.3 Template Matching

The authors propose the template matching technique based on the direction of gradients to match object template images, in Figure 2. The method measures the similarity between the template (T) and the live image of target (I). The authors utilize the default implementation in the OpenCV library, which consists of two pyramid levels and sixty-three features.

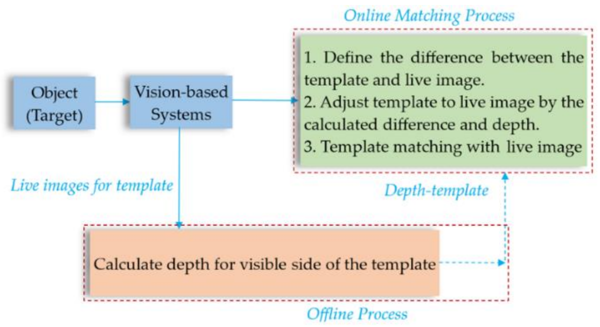


Fig. 2: Proposed template matching for vision-based abnormal object detection system

The measured similarity is expressed by the minimum distortion or maximum correlation position between T and I. Furthermore, the measured distortion is equal to the sum of squared differences in (2) as follows:

$$J(d)=\sum_{q_i}\left(I(q_i+d)-T(q_i)\right)^2 \qquad (2)$$

where presents the template's positions (pixel coordinates) and N presents the number of positions. Instead of producing them from CAD (Computer-Aided Design) models, the authors obtained offline several templates per object in various postures using a fixed camera. Additionally, the authors augmented the data to increase the number of abnormal templates. For each template, the authors included a 90% and 110% scaled variant. As our experimental process of a robotic work cell, where objects are believed to be put on a flat surface at a fixed, suitable adjusted distance from the camera, this was deemed sufficient. Items are matched at runtime, beginning with the highest ranked objects from our preprocessing. If no template from a given object has a score greater than a predetermined threshold, the next object on the list is examined. This procedure is repeated until a template's matching exceeds the threshold score. This template is then regarded matched, and the algorithm for matching templates is terminated. Importantly, we used fewer templates than the original implementation of. In addition, the authors employed higher resolutions 980x720 images replacing medium resolutions 640x480 images.

2.4 Background subtraction

Background subtraction provides a reasonable balance between the computational scale time and the detection rate. Figure 4 illustrates the background subtraction following procedure steps:

Step 1: Background initialization includes the techniques such as follows: back-ground creation, background extraction, and backdrop reconstruction, in the computing process of the initial background.

Step 2: Background representation explains the mathematical model used to illustrate the background.

Step 3: Background maintenance refers to the technique for updating the model so that it can adapt to changes that occur over time.

Step 4: Pixels' classification of the background and moving objects to detect whether foreground is the "background" or "moving objects" class.

The process of background subtraction comprising the subsequent steps (Figure 3):

- The background initialization will generate the first background based on the number of training frames N.
- The foreground detection classifies pixels into the "background" or "moving objects" class by comparing the background to the live images.
- The background maintenance will continuously refresh the background image by using all four above steps.

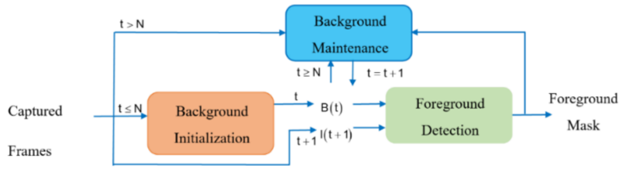


Fig. 3: Background subtraction process where N is the number of frames used for the background initialization and Bt is the background image at t, whereas It is the live image

The background representation combines with background maintenance have been conducted repeatedly as time progresses. Two confirmed images are chosen to be compared. Firstly, the sub-entity in background image is compared with sub-entity in the current image to obtain the pixel, region, or cluster dimensions. Next, the sub-entity is presented by a "feature" immediately received by the sensors with color features and stereo features). In addition, they are enhanced processing by handcrafted with edge features, texture features and motion features. Finally, AI camera for Deep Learning (DL) and deep data processing are used for optimal process.

3 RESULTS AND DISCUSSION

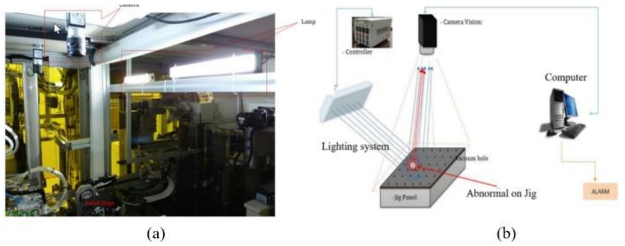


Fig. 4: Experimental system with (a): vision-based panel quality monitoring system and (b): main components

First, with a proposed object detection network, the pretraining image database is taken based on more than 15,000 images containing small objects in COCO's dataset [Lin 2014]. Next, the automatic vision system is equipped with camera: Basler acA2500-14gm; Len: TV1614-5MP; LED: CDL-200 with Light controller: CDP-4CH (Figure 4). Next, the operation diagram is shown as in Figure 5. The suitable devices of the automatic vision-based system are arranged in Figure 5c.

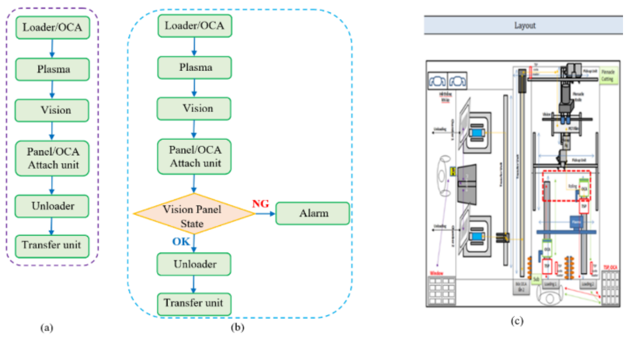


Fig. 5: The vision-based panel quality monitoring system with (a): the flow chart of the automatic vision process

without monitoring, (b): under monitoring, and the layout of automatic system

Based on an automatic mechanical system, the vision program will be created to identify the abnormal in the test mode with following required technical specifications (Figure 6).

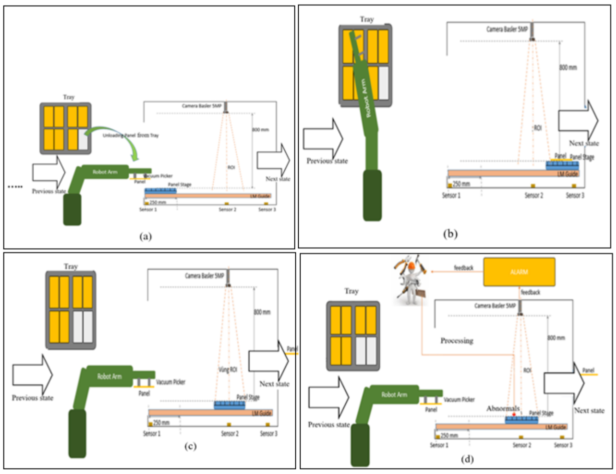


Fig. 6: The operation steps of automatic vision-based quality monitor system with (a) Step 1: Loading panels; (b) Step 2: Plasma surface treatment; (c) Step 3: Set up the vision area; and (d) Step 4: Vision-based checking process

After four steps of an automatic mechanical system have been completed, the vision process will be carried out sequentially (Figure 6). First, Canny Edge Detection is utilized to preprocess the main edges of the collected mages. Then, a Gaussian kernel removes the image's noise. Next, the Sobel kernel produces a pre-liminary edge image. All image's pixels are checked to create the edges, being only one pixel is a local maximum. Furthermore, hysteresis thresholding is used for removing the pixel which is not edged. Canny edge detection with identifying end-points algorithm will connect discontinued edges. Finally, rectify the discrepancy between the peaks of the original image's left and right edges and the live images. After completing the correction process of the deviation between live images and the original image, the authors utilize template matching method to register the state of panel (OK). All features in OK panel are defined completely with the contour, vacuum hole, scratch, etc. (Figure 7).

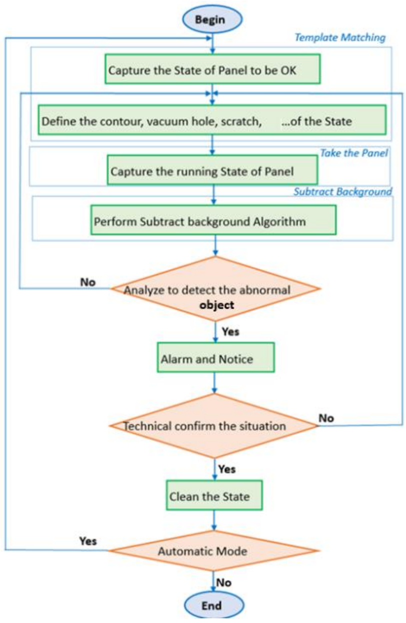


Fig. 7: The flow chart of monitoring process

Finally, after checking successfully with the test mode, Figure 8 continuously employs a subtraction background and template matching technique for the detection of minor abnormal objects with actual images of panel in production. Four steps of the augmentation method include defining the original coordinate, the derivation, rotation and calculation parameters after moving.

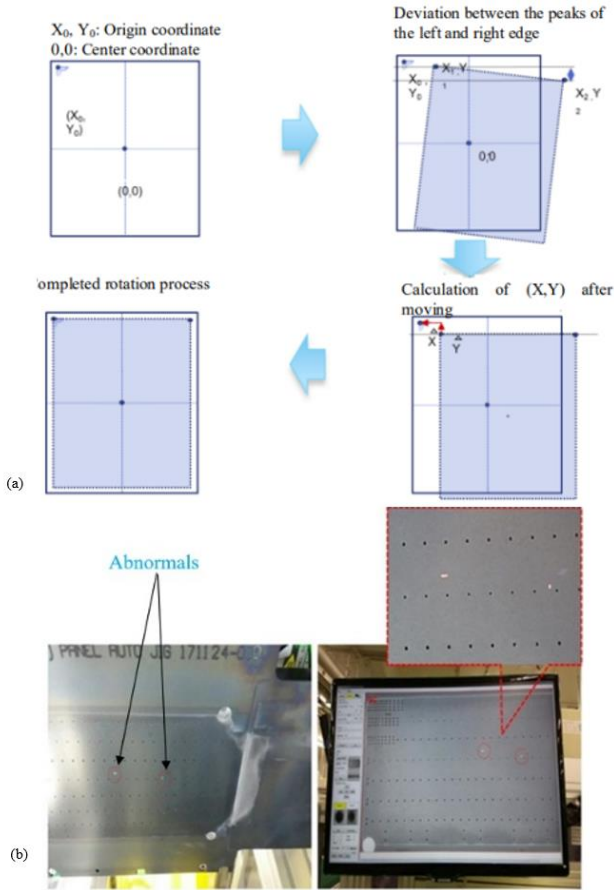


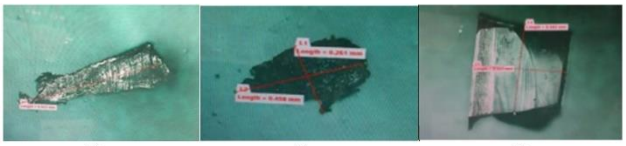
Fig. 8: Abnormal detection with (a) four steps of the augmentation method and (b) real abnormal

Through experimental processing and evaluating with the same trained data set, the proposed abnormal detection model is compared with a template matching method combining with normalized cross correlation techniques (NCC) [Amri 2020]. Our method shows high accuracy and fast FPS in Table 1. Moreover, the proposed method using background subtraction increases the detection performance.

Tab. 1: Small abnormal object detection results in the comparison with different template matching methods

Model	Acc (%)	FPS
The template matching method without NCC techniques [Amri 2020]	78÷82	17÷20
The template matching method having NCC techniques [Amri 2020]	85	19÷21
Our proposed method	**92**	**21÷25**

Furthermore, in the 10 days of continuous production, the abnormal object was detected and captured the image at the test panel of position 3, 28 and 50. The size of detected abnormal is 0.92 mm x 0.4 mm, 0.45 mm x 0.26 mm, and 0.5 mm x 0.5 mm, respectively (Figure 9).



(a) (b) (c)

Fig. 9: Actual abnormal are successfully detected obtained at (a): panel's position 3; (b): panel's position 28; and (c) panel's position 50

In addition, the automatic vision procedure occurred in 10 consecutive days with 19054 Panel items. The average accuracy in Table 2 is approximately 91 %. Compared to two levels of quality assurance (QA): no management and naked eyes, the Not Good (NG) rate reduced from 38 % and 19 %, respectively to 3 % when an automatic vision technique was used.

Tab. 2: The accuracy of vision process in 10 continuous days with 19054 products of panel

Day	Number of Panels	(NG) Panels	Confirmation results	Acc (%)
1	1568	15	13	87.7
2	2314	22	19	87.4
3	2136	21	18	87.2
4	1997	13	12	92.3
5	2712	19	16	86.8
6	1365	7	7	100
7	2001	29	27	93.1
8	2452	31	28	90.3
9	965	12	11	91.7
10	1544	13	10	89.4
Total	**19054**	**182**	**161**	**88.5**

4 CONCLUSIONS

The paper put forth a methodology for monitoring small anomalous objects using a DL algorithm-based object detection network. Also, background elimination is incorporated into the DL network alongside template matching. The obtained anomalous images will be utilized as training data for the object detection and classification model, utilizing the updated dataset of object images. A decline in the NG rate to 3% has been accompanied by a 91% average precision. Furthermore, it is ensured that the tact time of each device is below 5 seconds. The efficacy of the suggested approach is substantiated by the tangible results. The vision-based technique has been improved in terms of accuracy and efficiency through the acquisition of anomalous images for the purpose of updating the data set. The proposed small anomalous detection network model will be improved in subsequent iterations to accommodate a greater number of abnormal classes with the data set always being refreshed by physical samples collected during the production process. Finally, automatic panel quality monitoring system raises its accuracy based on the abnormal image obtained in practical manufacturing.

5 ACKNOWLEDGMENTS

This research is funded by Hanoi University of Science and Technology (HUST) under project number T2023-PC-008. Lab-506: Computer Vision and Autonomous Mobile Robot (CVAMS), VJIIST, HUST, Vietnam and Shibaura Institute of Technology, Japan are gratefully acknowledged for providing work location, guidance, and expertise.

6 REFERENCES

[Amri 2020] Amri, I.K., and Murinto, K.A. Detection of Object Changes in Low Contrast Image Using Template Matching Method. IOP Conf. Series: Materials Science and Engineering, 2nd International Conference on Engineering and Applied Sciences (2nd InCEAS), 2020, Vol. 771, No. 1, pp. 012021. https://doi.org/10.1088/1757-899X/771/1/012021

[Bukala 2020] Bukala, A. et al. A Study on Pattern Recognition with the Histograms of Oriented Gradients in Distorted and Noisy Images. Journal of Universal Computer Science, 2020, Vol.26, No.4, pp. 454-478. https://doi.org/10.3897/jucs.2020.024

[Dang 2022] Dang, T.V. Smart home Management System with Face Recognition Based on ArcFace Model in Deep Convolutional Neural Network. Journal of Robotics and Control, 2022, Vol.3, No.6, pp. 754–761. https://doi.org/10.18196/jrc.v3i6.15978

[Dang 2023a] Dang, T.V. Smart Attendance System based on Improved Facial Recognition. Journal of Robotics and Control, 2023, Vol.4, No.1, pp. 46–53. https://doi.org/10.18196/jrc.v4i1.16808

[Dang 2023b] Dang, T.V. and Bui, N.T. Multi-Scale Fully Convolutional Network-Based Semantic Segmentation for Mobile Robot Navigation. Electronics, 2023, Vol.12, No.3, pp 533. https://doi.org/10.3390/electronics12030533

[Dang 2023c] Dang, T.V. and Bui, N.T. Obstacle Avoidance Strategy for Mobile Robot Based on Monocular Camera. Electronics, 2023, Vol.12, No.8., pp 1932. https://doi.org/10.3390/electronics12081932

[Dang 2023d] Dang, T.V. et al. IRDC-Net: Lightweight Semantic Segmentation Network Based on Monocular Camera for Mobile Robot Navigation. Sensors, 2023, Vol. 23, No. 15, pp. 6907. https://doi.org/10.3390/s23156907

[Dou 2018] Dou, J. et al. Robust image matching based on the information of SIFT. Optik, 2018, Vol.171, pp. 850-861. https://doi.org/10.1016/j.ijleo.2018.06.094

[He 2015] He, K. et al. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1026- https://doi.org/1034. 10.1109/ICCV.2015.123

[Lin 2014] Lin, M.M. et al. Microsoft coco: Common objects in context. arXiv preprint, 2014, arXiv:1405.0312. https://doi.org/10.48550/arXiv.1405.0312

[Lin 2022] Lin, A., and Lu, A. Incorporating knowledge of plates in batch normalization improves generalization of deep learning for microscopy images. Proceedings of the 17th Machine Learning in Computational Biology meeting (PMLR), 2022, Vol. 200, pp. 74-93.

[Liu 2015] Liu, W. et al. SSD: Single shot multibox detector. arXiv preprint, 2015, arXiv:1512.02325. https://doi.org/10.48550/arXiv.1512.02325

[Mahaur 2023] Mahaur, B. et al. Improved Residual Network based on norm-preservation for visual recognition. International Neural Network Society, 2023, Vol.157, No.3, pp. 305-322. https://doi.org/10.1016/j.neunet.2022.10.023

[Nguyen 2023] Nguyen, V.T., and Chu, D.T. Study on Tracking Real-Time Target Human Using Deep Learning for High Accuracy. Journal of Robotics, 2023, Vol.2023, No. 9446956. https://doi.org/10.1155/2023/9446956

[Oron 2017] Oron, S. et al. Best-buddies similarity-Robust template matching using mutual nearest neighbors. IEEE Transactions on Pattern Analysis and Machine Intelligent, 2017, Vol.40, No.24, pp. 1799-1813. https://doi.org/10.1109/TPAMI.2017.2737424

[Tang 2019] Tang, G. et al. Distinctive image features from illumination and scale invariant keypoints. Multimedia Tools and Applications, 2019, Vol.78, No.01, pp. 23415–23442. https://doi.org/10.1007/s11042-019-7566-8

[Youngmo 2021] Youngmo, H. Reliable Template Matching for Image Detection in Vision Sensor Systems. Sensor, 2021, Vol.21, No.24, pp. 8176. https://doi.org/10.3390/s21248176