UDS 004

RESEARCH ON KEY TECHNOLOGIES OF EMBEDDED ARTIFICIAL INTELLIGENCE FOR MACHINE VISION

Sun Cuigai, Master of Engineering, Associate Professor, Department of Big Data and Internet, Suzhou College of Information Technology
（Suzhou City, Jiangsu Province, China）
Anna Tailakova, Scientific supervisor, Candidate of Technical Sciences, Associate Professor of the Department of Applied Information Technologies
（ T. F. Gorbachev Kuzbass State Technical University）
Wang Yihuai, Doctoral Supervisor, Professor, Head of the Department of Computer Science and Technology, Suzhou University
（Suzhou City, Jiangsu Province, China）

Abstract: This paper focuses on the research of key embedded artificial intelligence technologies for machine vision, conducting a series of studies from key technology analysis and system architecture to application practice. It constructs a complete embedded AI application development system called AHL-EORS. The system consists of two main components: the PC-side software (EORS) and the terminal system (AHL-D1-H). EORS provides full-process development support, including data collection and annotation, model training and testing, and component generation. Finally, the practical value of the AHL-EORS system is verified through an example of surface defect detection in welding pieces.

Keywords: Machine Vision, Embedded Systems, Artificial Intelligence, Defect Detection

1 Introduction

With the rapid development of Internet of Things (IoT) technology and smart devices, embedded artificial intelligence (AI) technology has been widely applied in various fields, especially in the field of machine vision. Due to their small size, low power consumption, and strong real-time performance, embedded AI systems are transforming traditional image processing and pattern recognition methods. From facial recognition on smartphones to quality inspection in industrial automation, from environmental perception in autonomous driving to behavior analysis in smart homes, the integration of embedded artificial intelligence and machine vision is creating significant social and economic value. This research aims to explore key technologies of embedded artificial intelligence for machine vision in depth, propose innovative solutions, and combine embedded systems, IoT, machine vision, and artificial intelligence through application practices such as defect detection in welding pieces. These solutions feature informatization, digitalization, and intelligence.

2 Key Technology Research

2.1 Hardware Architecture and Key Peripheral Components

The hardware architecture of the GEC system is built around the Allwinner

D1-H chip, incorporating a modular design and supporting hardware resources, named AHL-D1-H. It includes the D1-H chip and its minimal hardware system, SPI Nand Flash, tri-color LED, reset button, and two TTL-USB serial ports. The key peripherals involved in the terminal template mainly include the LCD module. Among these, the camera module is responsible for image acquisition, providing input data for the inference model; the LCD module is used to display images, inference results, and other critical information, helping users intuitively understand the system's operating status. As shown in Figure 1.

Figure 1 Hardware Architecture and Key Peripheral Components

2.2 Architecture Design of Terminal Inference Template

The terminal inference template project is based on the self-developed GEC (General Embedded Computer) system, providing users with a predefined software framework that completes the basic structure and functional modules for specific tasks via GEC. By leveraging the EORS software to automatically generate and replace inference components, it enables users to quickly customize and deploy applications according to different needs. To enhance the reusability and portability of embedded projects, the terminal inference template adopts component-based technology, dividing the entire project's functionality into several independent and replaceable components. Users can modify model types, adjust model parameters, and define new image classification task labels based on actual requirements.

2.3 Deep Learning Model Training and Optimization

In the EORS software, the training and optimization of deep learning models involve multiple stages, including data preprocessing, model construction, loss function selection, optimizer selection, and hyperparameter tuning. EORS uses stratified sampling through random partitioning from machine learning model libraries to ensure that the training and test sets maintain consistent category proportions. The input image size is determined based on the requirements of the network architecture and the actual application. EORS employs lightweight deep learning models to ensure classification accuracy while reducing computational resource consumption and improving inference speed. For loss function selection, EORS primarily uses cross-entropy loss functions based on the task type and model performance requirements, along with the Adam optimizer. Additionally, EORS provides options for hyperparameter tuning, allowing users to select appropriate training configurations based on task requirements and computational resource constraints, including learning rate, batch size, and number of training epochs.

3 Welding Defect Detection Application Practice

3.1 Overall System Framework

Traditional methods for inspecting the appearance of welding components rely on manual visual inspection, which is inefficient and prone to errors. Therefore, a machine vision-based approach is adopted, leveraging the self-developed AHL-EORS Welding Appearance Defect Detection System to improve the accuracy and efficiency of welding appearance inspections, enabling intelligent management and control of welding quality.

The system follows the design principle of "training on the PC and inference on the terminal" and is divided into four parts: image acquisition, model training, terminal deployment, and terminal inference. The image acquisition module uses a camera to capture images of the welding component's appearance. These images are labeled to form a dataset of welding spot samples, which is then input into the PC for model training. The trained model parameters are converted into a format that can be directly compiled and used as cognitive model parameter components within a general embedded engineering framework. Finally, the cognitive model component is deployed on the terminal, enabling the terminal to perform welding appearance defect detection. The architecture of the welding appearance defect detection system is shown in Figure 2.

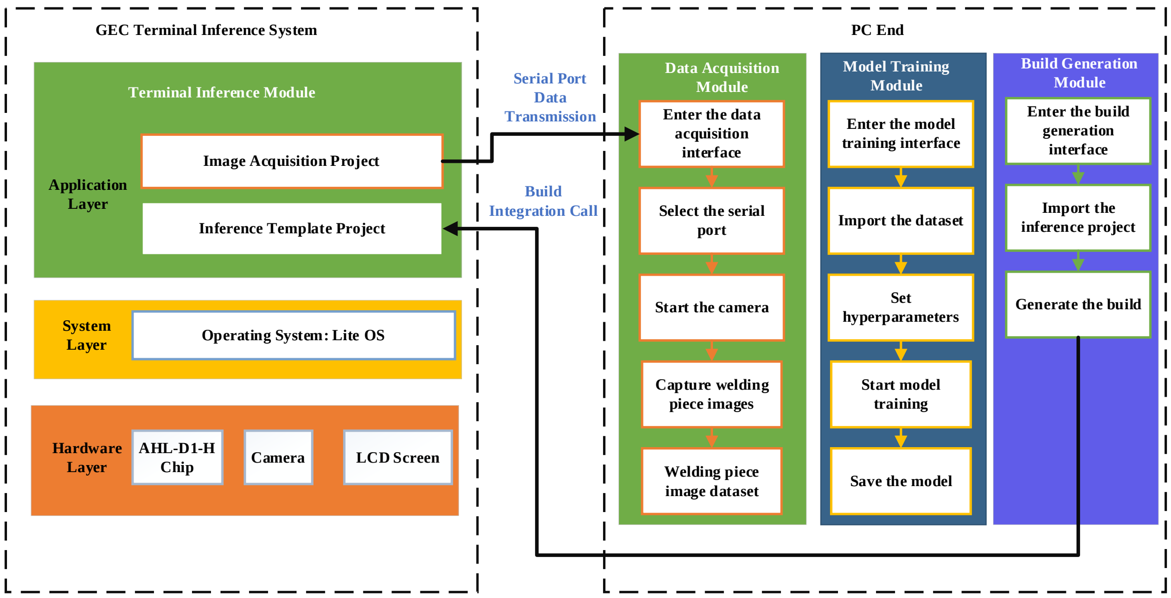
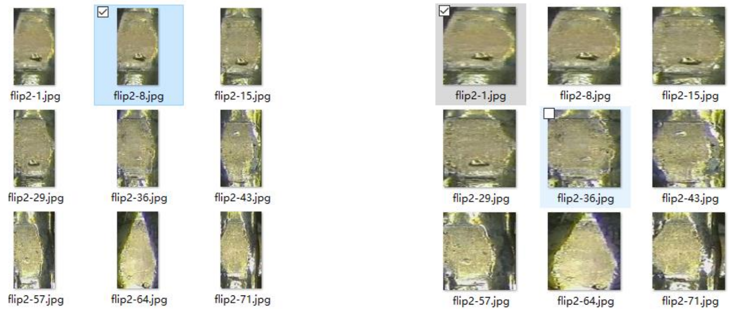


Figure 2 Architecture of the Welding Appearance Defect Detection System

3.2 Welding Component Recognition Model Training and Testing

(1) Data Preprocessing: The sampled image data has a pixel matrix structure of 160×85, while the standard training image structure for the AHL-EORS training model is 150×160 pixels. Therefore, batch conversion can be performed using specialized graphic conversion software, as shown in Figure 3.

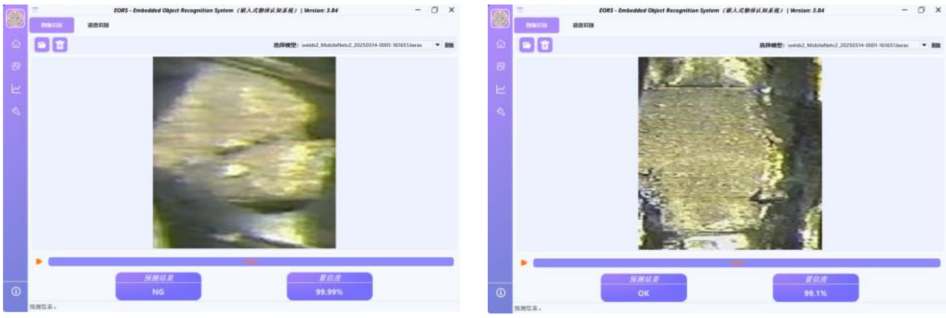


(a) 160×85 size (b) 150×160 size

Figure 3 Training Image Data Format Conversion

(2) Model Training: The prepared datasets of NG (defective) and OK (non-defective) images are imported into the training software, and the training parameters are set. The training is conducted over 40 epochs with a batch size of 16, a learning rate of 0.001, and a validation set ratio of 0.25.

(3) PC-Side Model Validation: The test images are imported into the validation page, and after selecting the trained model, the training results are tested, as shown in Figure 5.



(a) NG (b) OK

(4) Training Model Optimization and Hyperparameter Tuning: Optimizing and tuning the training model is a crucial step in the model training process. By adjusting the model's parameters, the training process and final results can be optimized. The goal of hyperparameter tuning is to find the best combination of parameters so that the model achieves optimal performance on the validation set while maintaining good generalization capabilities. In the case of welding appearance defect detection, the tuning process involves training with a fixed learning rate and evaluating the training effect based on the accuracy of the validation set. Four different learning rates were chosen for model training: 0.001, 0.01, 0.02, and 0.04. The training process is shown in Figure 4.

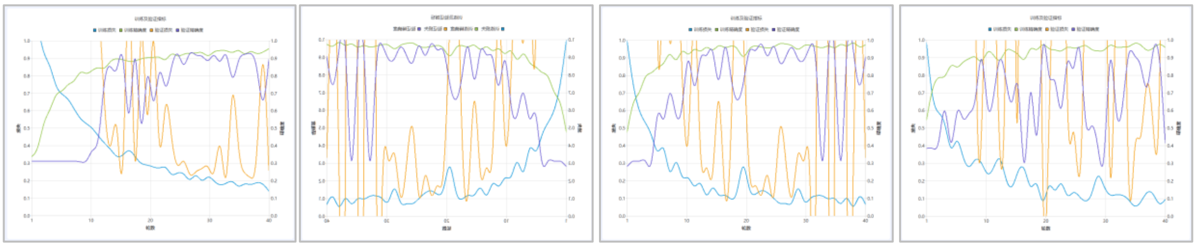


Figure4: Training Process of Welding Appearance Detection with Different Learning Rates

The validation accuracy of the test set after the training process for welding appearance detection with different learning rates is shown in Table 1

Table 1 Accuracy of Welding Appearance Inspection Validation Set with Different Learning Rates

Model ID	Learning Rate	Validation Set Accuracy (%)	Model File (with .keras suffix)
1	0.001	94.74	welds2_MobileNetv2_20250313-002233
2	0.01	88.88	welds2_MobileNetv2_20250312-221856
3	0.02	94.5	welds2_MobileNetv2_20250312-195422
4	0.04	50.64	welds2_MobileNetv2_20250312-230652

From Table 5, it can be seen that the overall test accuracy is highest when the learning rate is 0.001 and 0.02. The test set accuracy of the models trained with learning rates of 0.001 and 0.02 exceeds 94%, making them suitable for inference and recognition tasks.

(5) Component Generation: Using the component generation module of the EORS software, the welding defect detection model trained in the previous section can be converted into an executable inference component for the terminal and integrated into the terminal project. This is shown in Figure 5.

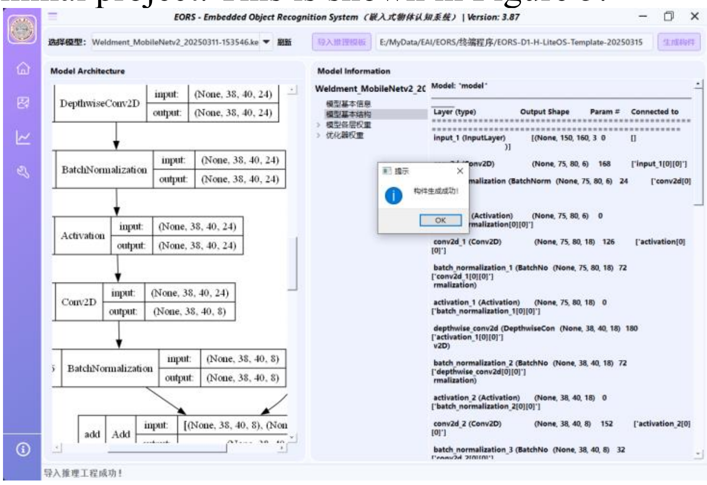


Figure5 Component Generation and Integration into the Terminal Project

(6) Model Deployment: The terminal inference framework utilizes the task scheduling mechanism of the LiteOS embedded real-time operating system. While ensuring the real-time performance of the core inference thread, a dedicated secondary programming thread is created. This thread can execute user-defined logic code based on the inference results. For the welding appearance defect detection task, the secondary programming thread function is implemented in the terminal inference project as follows:

```
void thread_secprg(void)
{
        uint32_t tEventState;
```

```
            while (1)
          {
              event_recv(g_EventWord,SECPRG_EVENT,\
EVENT_FLAG_OR|EVENT_FLAG_CLEAR,WAITING_FOREVER,&tEvent
State);
              if (g_img_pro_flag == 0)    goto thread_secprg_1;
              float ok_probability = g_result[0];
              float ng_probability = g_result[1];
              if(ok_probability > ng_probability){
                LCD_DrawSurface(36, 0, 186, 48, WHITE);
                LCD_DrawSurface(36, 208, 186, 256, WHITE);
              }else{
                LCD_DrawSurface(36, 0, 186, 48, RED);
                LCD_DrawSurface(36, 208, 186, 256, RED);
              }
              g_img_pro_flag = 0;
      thread_secprg_1:
              __asm("nop");
          }
        }
```

As can be seen from the implementation of the above code, the system achieves visual feedback of welding inspection results through secondary programming: when the welding inspection result is qualified (OK), the image background color is white to maintain a clean interface; when the welding appearance inspection result is unqualified (NG), the image background color is a high-contrast warning red, ensuring that quality inspectors can quickly and accurately identify defective workpieces.

3.3 Practical Application in Actual Production Lines

After completing laboratory testing of the weld inspection system, applying it to actual factory predictions is a critical step. Factory predictions can verify the feasibility and effectiveness of the weld inspection equipment in real production environments. The research team collaborated with TE Connectivity Technology Co., Ltd. to deploy this welding appearance quality inspection device to the resistance welding production line at the Suzhou TES factory, as shown in Figure 6.

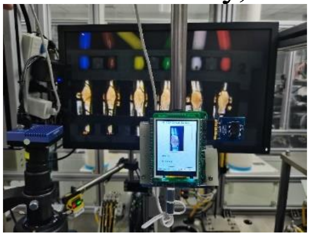
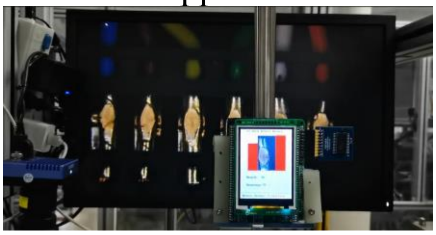
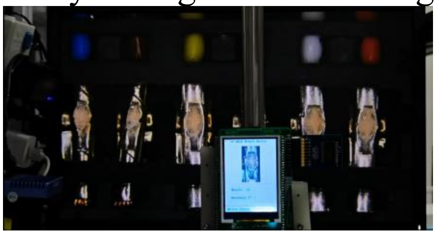


Figure 6 Factory Deployment

When an NG sample is detected, the display screen scrolls red on both sides to alert the operator of the welding defect. When the sample is OK, there is no alert. Both states display the current detection accuracy. The chief engineer of the production line not only affirmed the detection accuracy and efficiency of the equipment but also provided valuable experience and feedback for further optimization and improvement of the system. They commented that if this detection device were deployed in the resistance welding production line, it could significantly improve production efficiency and product quality, thereby promoting the optimization and improvement of the entire welding process. It has significant research significance and application value. Factory testing is shown in Figure 7.



(a) NG Detection



(b) OK Detection

Figure 7 Factory Testing

(8) Test Summary. To address the issue of appearance quality defects in electronic connector welding components, a complete intelligent inspection system was implemented using the AHL-EORS system. Trial operation results showed that the recognition accuracy rate for normal welds reached 95.5%, and the precision rates for various defect types were as follows: cracks at 90.6%, overheating at 100%, surface roughness at 91.3%, and weak welds at 100%. The system achieved an average detection speed of 3 seconds per weld point, comparable to manual inspection efficiency but with superior stability. Through a carefully designed human-machine interface, when an NG product is detected, the display screen flashes red on both sides in real time to warn operators, significantly improving their work efficiency. During a continuous operation test lasting 26 days, the system accumulated 58 hours of fault-free operation, fully demonstrating its stability and reliability.

4 Conclusions

This paper focuses on the key technologies of embedded artificial intelligence for machine vision and develops the embedded AI application development system AHL-EORS. At the theoretical level, it clarifies the related concepts of machine vision, embedded computing, and embedded artificial intelligence, and investigates core technologies such as embedded AI chips and model lightweighting. Technically, the PC software and terminal system of the AHL-EORS work collaboratively to provide full-process development support and an efficient real-time inference environment. Through practical application in welding appearance defect detection, the accuracy, stability, and reliability of the system were verified. The system demonstrated high detection precision and fast speed, gaining recognition in real-world applications. The research results not only provide an effective solution for the application of embedded artificial intelligence in the field of machine vision but also promote the development of related technologies, enhancing the level of production intelligence. With broad application prospects and significant practical value in industrial inspection and other fields, this study positively contributes to promoting the deep integration of embedded AI and machine vision technologies.

References

[1] Chen Chao, Zhou Xiangxue. Design of a commodity delivery robot based on machine vision technology[J]. Electronics Making. 2024, 32(05): 65-68.

[2] Dong Huiwen, Yu Bicheng. Design and implementation of an embedded-based visual tracking system[J]. Modern Computer (Professional Edition). 2016 (16): 36-38.

[3] Wu Maojun, Lin Yunfeng. Research on defect detection technology for wooden toy parts based on machine vision[J]. Automation & Instrumentation. 2025, 40(03): 86-91.

[4] Fu Sanli. Design of a real-time monitoring and statistical alarm system for pedestrian counting based on machine vision[J]. Plant Maintenance Engineering. 2023 (16): 152-153.

[5] Jia Yukun, Shen Shujun. Elevator internal defect monitoring and early warning system based on machine vision technology[J]. Modern Manufacturing Technology and Equipment. 2021, 57(02): 58-59.

Acknowledgements

2024 Provincial Education Science Planning Key Project "Research on the Practice Path of Digital Transformation of Vocational Colleges from the Perspective of New Quality Productivity" (Bb/2024/02/124); Excellent Teaching Team Project of 'Blue and Green Project' in Jiangsu Universities.

Brief Introduction of the First Author: Sun Cuigai, female, Associate Professor. Her research focuses on the teaching and research of courses such as mobile application development and artificial intelligence. email: hudie0011@163.com.