WEP Warwick Evans Publishing

International Journal of Mechanical and Electrical Engineering

ISSN: 3005-9615 (Print), ISSN: 3005-7132 (Online) | Volume 6, Number 2, Year 2025
DOI: https://doi.org/10.62051/ijmee.v6n2.03
Journal homepage: https://wepub.org/index.php/IJMEE/index

Research on Damage Detection and Recognition System for Automotive Components Based on Stereo Vision and Deep Learning

Yi Zhou and Haozhe Zhao *

School of Automotive Engineering, Wuhan University of Technology, Wuhan, Hubei, China
* Corresponding author: Haozhe Zhao (Email: 1910494992@qq.com)

ABSTRACT

Damage detection in automotive components is of paramount importance for ensuring vehicle safety and performance. However, traditional detection methods suffer from significant limitations in both efficiency and accuracy. The recent advancements in deep learning and stereo vision technologies have introduced innovative approaches for intelligent damage detection in complex scenarios. This study proposes a damage detection and recognition system for automotive components that integrates stereo vision and the YOLOv5 deep learning algorithm. The research methodology includes constructing a stereo vision data acquisition platform, conducting image preprocessing and depth information extraction, applying various data augmentation techniques to enhance sample diversity, and leveraging transfer learning and hyperparameter optimization to improve model performance. Experimental results demonstrate that the system exhibits excellent performance in detection accuracy, real-time capability, and adaptability to small objects, effectively identifying diverse types of component damage. This research provides a reliable technological foundation for intelligent detection tasks in complex industrial scenarios, contributing significantly to improving quality control efficiency.

KEYWORDS

Automotive Component Damage; Stereo Vision; Deep Learning; YOLOv5; Object Detection; System Design.

1. INTRODUCTION

Automotive components are essential for ensuring vehicle safety and performance, yet their operating environments are often complex and variable, making them prone to mechanical damage, fatigue, and wear. Therefore, effectively detecting component damage to safeguard vehicle operation has become a critical research focus in the modern automotive industry. Traditional methods for damage detection, primarily relying on manual inspection and conventional image processing techniques, exhibit clear limitations: manual inspection is labor-intensive and subjective, while conventional image processing methods struggle to accommodate the intricate and dynamic characteristics of component damage [1,2], particularly under variations in texture, lighting, and background interference. To address these challenges, intelligent detection techniques based on deep learning have been introduced, offering significant improvements in both accuracy and efficiency.

In recent years, stereo vision technology has matured in applications such as 3D reconstruction and object detection. By emulating the human visual mechanism, stereo vision leverages disparity principles to obtain depth information, effectively addressing the deficiencies of traditional methods in 3D feature extraction. Simultaneously, deep learning, as a cornerstone of artificial intelligence, demonstrates remarkable feature extraction capabilities and accuracy in object detection and classification tasks. Among these, the YOLO (You Only Look Once) series of algorithms is renowned for its speed and efficiency, making it a popular choice for real-time object detection. The YOLOv5 version further optimizes network architecture and training strategies, achieving a balance between performance and computational complexity, thus providing robust support for damage detection in complex scenarios.

Against this backdrop, this study aims to develop a damage detection and recognition system for automotive components by integrating stereo vision and deep learning technologies. The system employs stereo cameras for data acquisition, capturing high-precision 3D depth information, and utilizes YOLOv5 as the core detection algorithm to enable rapid and accurate surface damage identification[3]. The overall methodology of this research includes designing and constructing a stereo vision data acquisition platform for capturing and preprocessing images of various damage types, creating a dataset through annotation and augmentation techniques, conducting training, optimization, and deployment of the YOLOv5 algorithm, and finally, evaluating the system's performance in real-world application scenarios.

2. RELATED TECHNOLOGIES AND THEORETICAL FOUNDATIONS

2.1. Principles of Stereo Vision

Stereo vision is a computer vision technique that emulates the human visual system, capturing images of the same scene from two viewpoints using stereo cameras and calculating the depth information of objects based on geometric relationships. Renowned for its non-contact nature, efficiency, and relatively low cost, stereo vision has been widely applied in 3D reconstruction and scene understanding[4,5].

The fundamental principle of stereo vision lies in disparity, which refers to the positional difference of the same object in the left and right images. Disparity is a function of the distance between the object and the camera: the closer the object, the greater the disparity; conversely, the farther the object, the smaller the disparity. Based on geometric optics, stereo vision systems calculate object depth using known camera baseline distances (the fixed separation between the two cameras) and disparity values through triangulation. Specifically, the depth Z can be calculated as

$$\mathrm{Z} = \frac{f \cdot B}{d} \tag{1}$$

where Z represents depth, f is the camera focal length, B is the baseline distance, and d is the disparity. This formula underscores the dependency of depth information extraction on accurate disparity estimation, making disparity calculation a critical aspect of stereo vision technology.

In 3D reconstruction, stereo vision generates a comprehensive depth map by calculating the disparity value for each pixel in the scene. Depth maps provide an intuitive representation of spatial relationships among objects in a scene and serve as foundational data for further 3D point cloud reconstruction and model generation. For instance, in automotive component damage detection, depth information aids in isolating damaged areas from the background and precisely locating the damage, including its spatial dimensions. Compared to monocular vision, stereo vision offers a significant advantage by delivering both 2D image information and 3D structural data, enhancing robustness and accuracy in complex environments[6].

The application of stereo vision relies heavily on high-precision calibration and matching processes. First, geometric calibration of the stereo cameras is essential to obtain intrinsic and extrinsic parameters (e.g., focal length, distortion coefficients, baseline distance), ensuring the accuracy of depth computation. Second, stereo matching algorithms are used to identify correspondences of the same object between the left and right images. Commonly employed methods include block matching and feature-based algorithms, which optimize cost functions to improve disparity computation accuracy.

In summary, stereo vision technology achieves precise depth measurement through disparity calculation, forming the basis for 3D reconstruction. In this study, stereo vision not only provides training data enriched with depth information but also aids in extracting 3D features during damage detection, thereby improving accuracy and reliability. This capability of 3D reconstruction based on disparity and depth computation significantly enhances the overall performance of deep learning models.

2.2. Overview of YOLOv5 Algorithm

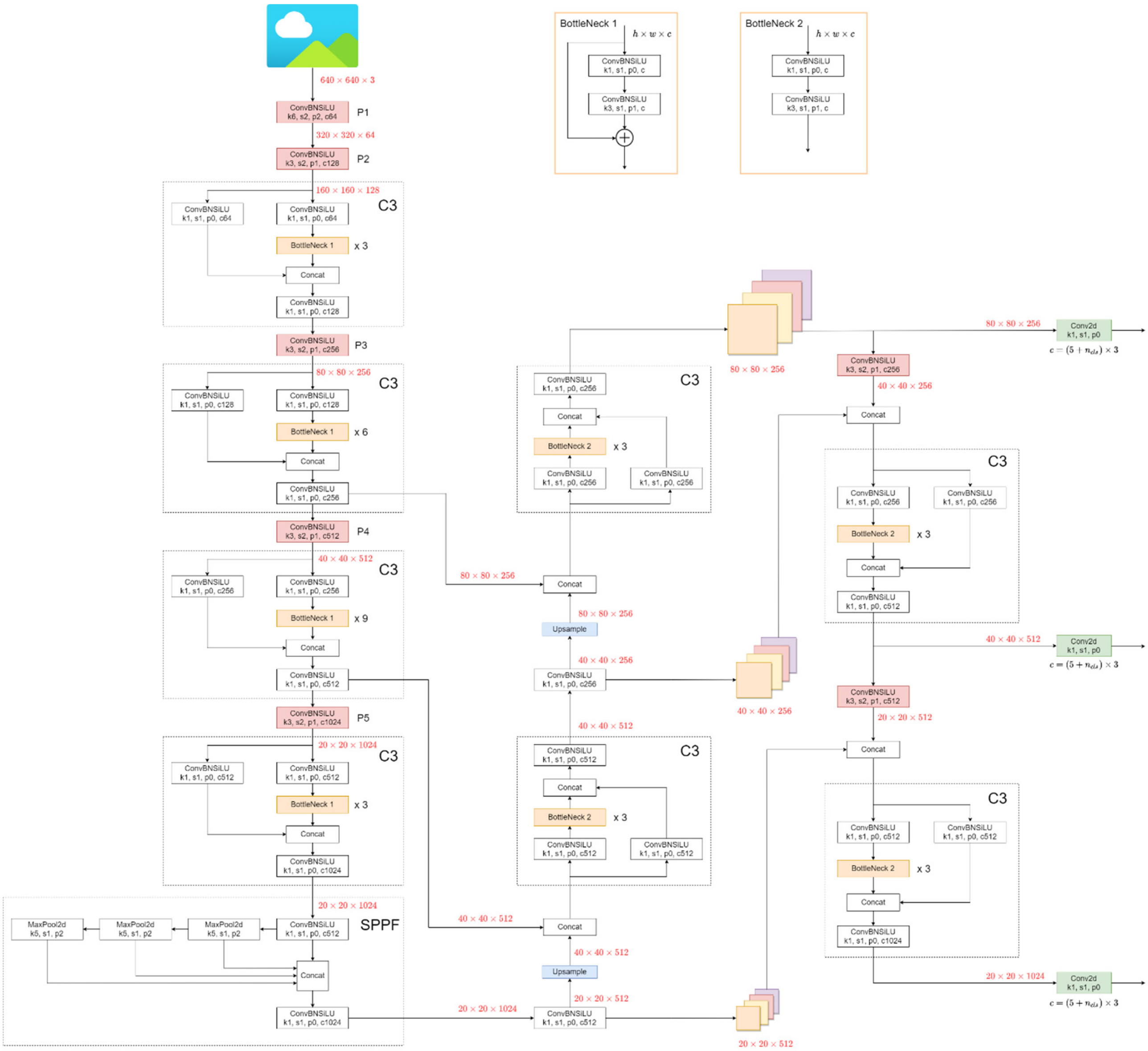

Fig 1. YOLOv5 Architecture

YOLOv5 (You Only Look Once, version 5) is a prominent version within the YOLO series of object detection algorithms, renowned for its high detection speed and exceptional accuracy. Building upon the core philosophy of YOLO, YOLOv5 introduces architectural improvements and optimizations, making it highly adaptable across diverse application scenarios [7].

The architecture of YOLOv5 is grounded in a single-stage object detection framework, differing from traditional two-stage detection algorithms like Faster R-CNN, as it completes both object localization and classification in a single forward pass, significantly enhancing detection efficiency. YOLOv5 comprises four primary modules: the Input Module, Backbone, Neck, and Head. The Input Module employs data augmentation and preprocessing techniques to improve model robustness against variations in object scale and orientation. The Backbone, featuring a CSPDarknet structure, effectively extracts multi-layered feature information. The Neck utilizes a combination of FPN (Feature Pyramid Network) and PANet (Path Aggregation Network) to strengthen multi-scale feature representation, while the Head employs adaptive bounding box regression and classification methods to output object locations and categories. The architecture of YOLOv5 is illustrated in Figure 1.

Compared to earlier YOLO versions, YOLOv5 exhibits notable advantages. First, its lightweight design and faster inference speed make it particularly suitable for deployment on resource-constrained embedded devices or real-time systems. Second, YOLOv5 incorporates advanced techniques such as Mosaic data augmentation, CIoU loss functions, and adaptive anchor generation to further enhance detection accuracy and adaptability to complex backgrounds. Additionally, YOLOv5 provides various model sizes, ranging from Nano to Xlarge, allowing users to flexibly balance speed and accuracy according to specific application needs.

The application of YOLOv5 in object detection leverages the automatic feature extraction and learning capabilities of deep learning models. Initially, the convolutional neural network (CNN) extracts feature maps from input images, capturing the spatial distribution characteristics of objects. Subsequently, the feature fusion network transmits cross-layer information, improving the detection of small objects and complex backgrounds. Finally, the prediction head generates results, including class probabilities and bounding box coordinates, while employing non-maximum suppression (NMS) to eliminate duplicate boxes and ensure the uniqueness and accuracy of detection outputs.

In this study, YOLOv5 serves as the core technology for detecting and recognizing damage in automotive components, offering distinct advantages. First, YOLOv5 can process stereo image data in real time, efficiently identifying component damage even in complex industrial environments. Second, its multi-scale detection capabilities and strong adaptability to small objects make it particularly effective in detecting subtle damage areas, such as scratches and cracks. Furthermore, through hyperparameter optimization and transfer learning, YOLOv5 achieves enhanced performance, ensuring detection stability and reliability across diverse scenarios.

3. SYSTEM DESIGN AND IMPLEMENTATION

3.1. Overall System Architecture Design

The damage detection and recognition system for automotive components proposed in this study adopts a modular design approach. The system architecture comprises four main modules: the data acquisition module, data preprocessing module, deep learning model training and optimization module, and model inference and application module. These modules work collaboratively to form a comprehensive end-to-end detection system, enabling fully automated processing from data collection to damage detection output.

The data acquisition module serves as the foundation of the entire system, responsible for capturing high-quality image data through stereo cameras. It includes hardware components, such as stereo cameras and their support structures, and control programs for data collection. By leveraging disparity principles, the stereo cameras synchronize the capture of left and right views, obtaining both 2D image information and depth data for the target object. During the acquisition process, techniques such as ambient light compensation and camera calibration are employed to ensure data quality and consistency, providing a reliable basis for subsequent processing.

The data preprocessing module focuses on data cleaning and optimization, encompassing tasks such as image correction, noise removal, and depth information extraction. This module integrates stereo image alignment algorithms based on image registration to eliminate distortion errors caused by camera disparity through geometric correction. Additionally, various image enhancement techniques, including color balancing and edge optimization, are implemented to improve image clarity and contrast, ensuring efficient feature extraction by the deep learning model.

The deep learning model training and optimization module constitutes the core of the system. It leverages the YOLOv5 algorithm to build, train, and optimize the target detection model. This module begins by precisely annotating damage regions in collected component images using data labeling tools and expands the dataset with data augmentation techniques to enhance the model's generalization capabilities. The YOLOv5 algorithm is then fine-tuned through transfer learning and hyperparameter optimization to adapt the model weights and architecture to various types of component damage. After training, the model is encapsulated and deployed within the inference environment.

Finally, the model inference and application module applies the optimized detection model to real-world scenarios, enabling real-time damage detection and recognition. This module integrates a model inference engine that generates bounding boxes and classification results based on input images while using depth information for 3D localization of damage areas. Furthermore, the module supports integration with industrial information systems, facilitating feedback of detection results to production line control systems for quality inspection and maintenance decision-making. The overall system framework is illustrated in Figure 2.

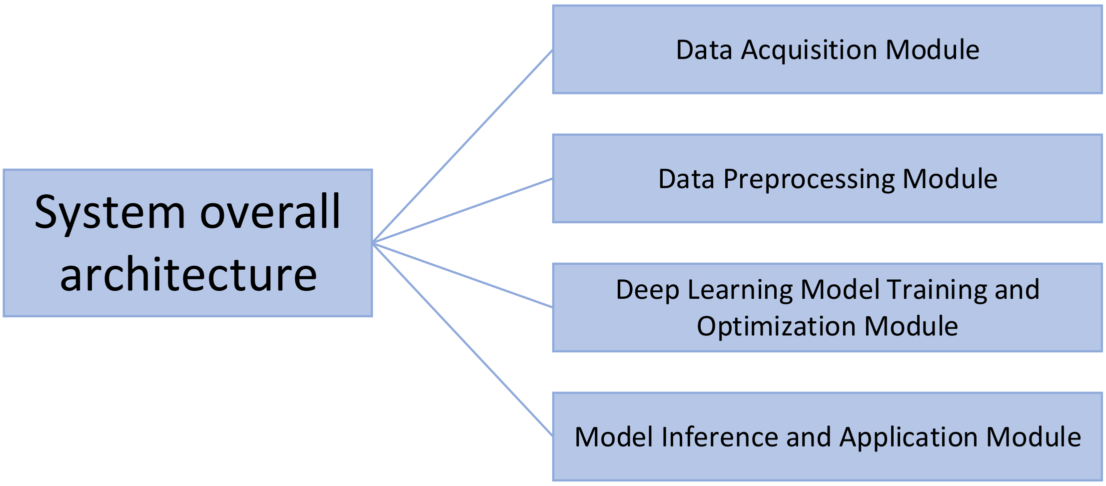

Fig 2. Overall System Architecture

The overall system architecture reflects the functional layering concept from the perception layer to the decision layer. The data acquisition and preprocessing modules constitute the perception layer, ensuring the accuracy and completeness of data. The deep learning model training and optimization module represents the analysis layer, delivering efficient detection algorithms, while the model inference and application module functions as the decision layer, responsible for interpreting and outputting detection results. In practical deployment, standardized interfaces enable seamless data exchange between modules, ensuring efficient system collaboration.

This system's architecture, centered on stereo vision and deep learning technologies, incorporates a modular and layered design philosophy, ensuring flexibility, robustness, and scalability. Through the efficient collaboration of the above modules, the system achieves precise and real-time detection of automotive component damage, providing critical technical support for intelligent manufacturing and vehicle maintenance.

3.2. Data Acquisition Module Design

The data acquisition module is a pivotal component of the system, tasked with capturing high-quality image data and depth information using stereo cameras. This module's design encompasses hardware selection, parameter configuration, hardware setup, and optimization of the data acquisition process to ensure the clarity, consistency, and integrity of the captured data.

This study employs the Intel RealSense D455 stereo camera, widely recognized for its precision and stability in 3D reconstruction and industrial inspection applications. Key features of the D455 include a baseline distance of 95 mm, which enables extended depth measurement ranges, with depth accuracy reaching ±2% at a distance of 2 meters. The camera provides a resolution of 1280×720 pixels and a maximum frame rate of 90 FPS, meeting the requirements for data acquisition in high-dynamic scenarios. Additionally, the camera's built-in Depth Processing Unit (DPU) delivers real-time depth maps, significantly reducing the computational load of post-processing. An illustration of the Intel RealSense D455 is shown in Figure 3.

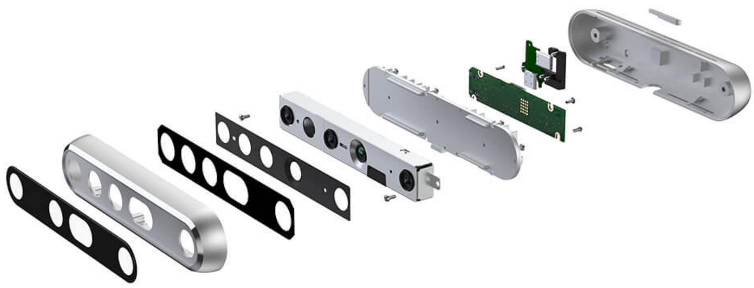

Fig 3. Intel RealSense D455 Illustration

To ensure data quality, key camera parameters such as focal length, aperture, and exposure time were optimized during configuration. The focal length was set to a standard value to balance image clarity and field of view. The aperture was configured in auto-adjustment mode to accommodate varying lighting conditions, while exposure time was manually adjusted based on experimental environments to prevent image detail loss due to overexposure or underexposure. Additionally, the camera's white balance function was activated to maintain consistent color representation across images.

To enhance the accuracy of depth information computation, the stereo camera underwent detailed geometric calibration during setup. Calibration involved capturing chessboard images to determine the camera's intrinsic and extrinsic parameters (e.g., focal length, principal point position, distortion coefficients) and generating a precise calibration matrix. These parameters were subsequently loaded into the data acquisition module to correct distortions encountered during the acquisition process.

The design of the data acquisition process considers multiple factors, including the acquisition environment, lighting conditions, and camera angles, to ensure the collection of high-quality training data. First, the acquisition was conducted in an indoor laboratory setting to minimize external interference. The laboratory was equipped with uniformly distributed LED lighting to eliminate uneven illumination that could cause variations in image brightness. Additionally, the background of the acquisition area was designed in neutral tones (gray or black) to reduce interference from non-target objects.

Lighting adjustments were a focal point of the data acquisition process. To ensure the clarity of surface details on components, the angle and intensity of light sources were carefully optimized.

Specifically, lights were positioned on either side of the target object at a 45° angle to highlight fine surface damage features. For highly reflective components, polarized filters were applied to reduce glare.

The configuration of camera angles and distances was tailored to the shapes and sizes of the components. Typically, the camera-to-object distance was set between 0.5 and 1 meter to ensure both image resolution and depth measurement accuracy. During the acquisition process, cameras captured frontal views from fixed angles while also employing multi-angle rotational shots to obtain comprehensive features of the target object, thereby enriching the diversity of data for subsequent model training.

After completing hardware setup and acquisition process design, the module was integrated into the system and optimized. Custom data acquisition software was developed to enable real-time adjustment of camera parameters and efficient storage of image data. The software supports batch data acquisition and is compatible with annotation tool interfaces, simplifying the data processing workflow. Furthermore, to improve acquisition efficiency and coverage, robotic arms were employed to automate camera positioning and angle adjustments, enabling comprehensive, seamless data capture.

3.3. Data Preprocessing

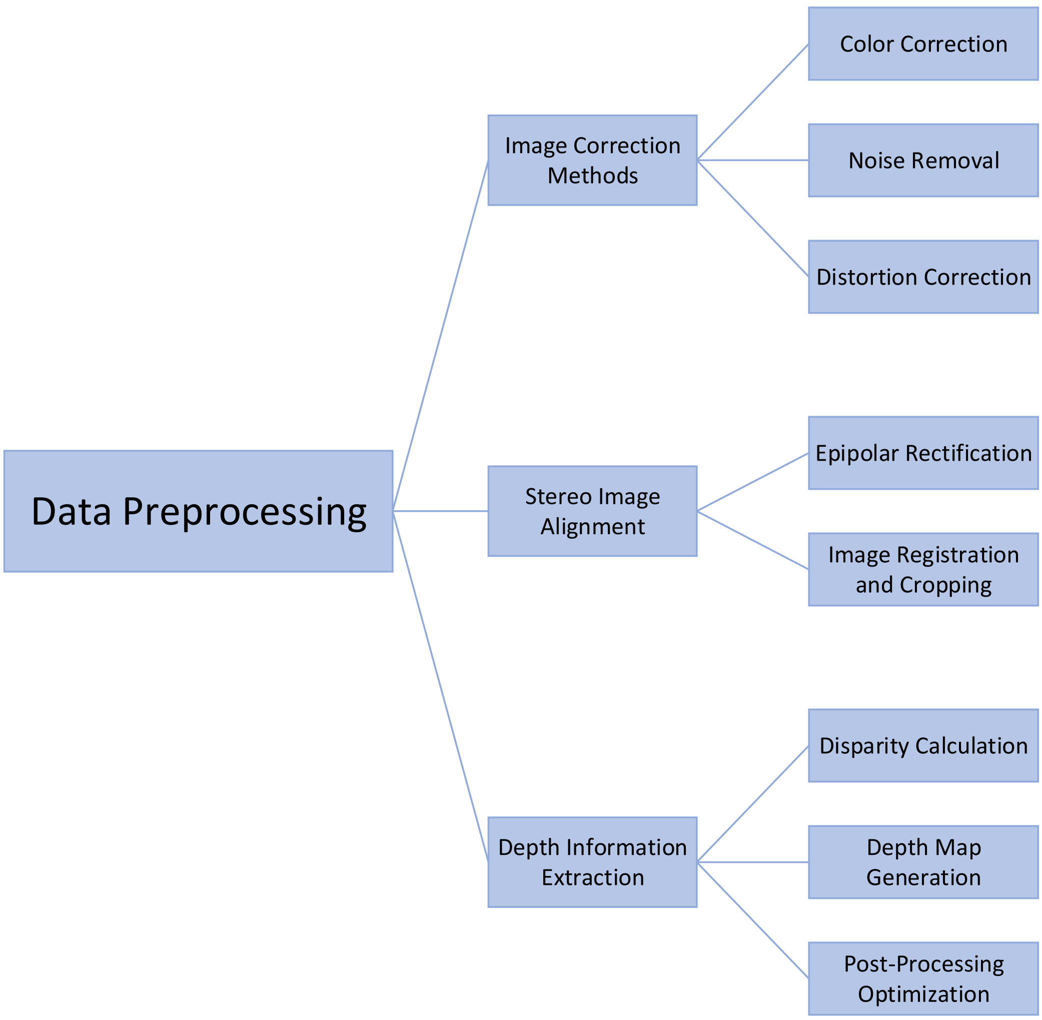

Fig 4. Data Preprocessing Methods

Data preprocessing is a critical step to ensure the subsequent performance of model training and inference. Its primary goal is to enhance the quality and consistency of raw data through various technical methods, thereby providing accurate and reliable input for deep learning models. In this study, data preprocessing focuses on key tasks such as image correction, stereo image alignment, and depth information extraction. The methodology is illustrated in Figure 4.

During image acquisition, factors such as lighting conditions and camera hardware characteristics may introduce color deviation or uneven distribution in the images. To address this, the system employs color balancing algorithms to correct image color and ensure uniformity. Specifically, white balance correction based on the gray-world assumption is used to mitigate color bias caused by varying light sources, while gamma correction enhances image brightness and contrast through nonlinear transformations. Additionally, for images of components with highly reflective surfaces, polarized filter correction is applied to reduce glare interference.

Raw images often contain noise from optical sensors or environmental backgrounds, which can affect the model's feature extraction accuracy. This study applies denoising methods based on Non-Local Means (NLM) and combines them with Gaussian filtering to smooth low-frequency noise. These techniques effectively preserve edges and texture details while eliminating redundant information, thereby improving data quality.

Due to the optical characteristics of camera lenses, acquired images may exhibit radial and tangential distortions. Using intrinsic parameters and distortion coefficients obtained during camera calibration, the system adopts OpenCV-based distortion correction algorithms to rectify these images. Specifically, pixel coordinates are remapped to undistorted coordinates through inverse mapping, restoring the geometric integrity of the images and ensuring the accuracy of subsequent stereo alignment and depth calculations.

The key to stereo vision is the precise alignment of the left and right views, which directly impacts the accuracy of disparity calculation and depth extraction. To achieve high-quality alignment, this study follows these steps: After stereo camera calibration, vertical alignment errors between the left and right images are common. Epipolar rectification is used to map corresponding points in both images onto the same horizontal line, eliminating vertical misalignment. The system also automatically crops aligned images to remove invalid regions created during rectification. Sub-pixel precision registration further optimizes alignment, ensuring the stability of disparity calculations.

Depth information extraction is the core task of stereo vision systems and involves three main stages: disparity calculation, depth map generation, and post-processing optimization. Disparity calculation forms the foundation of depth information extraction. The system employs cost-aggregation-based stereo matching algorithms, which calculate the matching cost for image blocks and use sliding window aggregation to enhance robustness. Dynamic Programming (DP) further optimizes disparity continuity, reducing noise and mismatches. Based on the geometric relationship between disparity and depth, disparity maps are converted into depth maps. To enhance depth map details, the system incorporates edge-preserving filters such as bilateral or guided filtering, smoothing depth maps while retaining object boundaries.

To further improve depth map quality, the system applies semantic-based hole filling techniques to address invalid regions (e.g., occlusion or reflection areas) in disparity calculations. Additionally, multi-scale fusion methods integrate depth maps at different resolutions to generate refined 3D structural information.

By implementing these preprocessing steps, the system significantly improves the quality and consistency of input data, providing optimized training samples for deep learning models. In complex industrial scenarios, precise image correction and depth extraction enhance the clarity of damage region features, thereby improving the model's ability to detect subtle damage. This module lays a solid technical foundation for achieving the overall performance objectives of the system.

3.4. Data Preprocessing

3.4.1. Data Augmentation and Processing

Data augmentation is an essential technique for improving the robustness and generalization of deep learning models. Particularly in object detection tasks, augmenting the diversity of training data can effectively mitigate overfitting issues and enhance the model's adaptability to real-world applications. In this study, various data augmentation methods are employed during YOLOv5 training to optimize the dataset across multiple dimensions, providing enriched input samples for the model.

Five commonly used data augmentation methods are implemented, including geometric transformations, color adjustments, noise processing, random cropping and padding, and Mosaic data augmentation.

Geometric transformations constitute a fundamental approach to data augmentation, simulating diverse spatial distributions of targets in real-world scenarios. Rotation randomly alters image orientation, enhancing the model's ability to recognize objects from different angles. Flipping (horizontal or vertical) is particularly effective for symmetric objects, increasing directional diversity within the dataset. Scaling adjusts object proportions to improve the model’s adaptability to varying distances in detection scenes. Translation changes object positions within the image, strengthening the model's robustness to spatial variations.

Color adjustments simulate variations in lighting conditions or capture devices, enhancing the model’s robustness to color changes. Brightness adjustments increase or decrease overall image luminosity, mimicking bright or dim environments. Contrast adjustments modify the differences between light and dark regions, sharpening object details and boundaries. Hue shifts randomly alter RGB channel values, enabling the model to handle color deviations under different lighting conditions. Gamma correction uses nonlinear transformations to adjust brightness distribution, amplifying features in darker or brighter regions.

Noise processing introduces or mitigates noise to simulate real-world challenges, enhancing the model's robustness. Gaussian noise introduces subtle pixel fluctuations, maintaining detection performance under reduced image quality. Salt-and-pepper noise adds random black or white pixels to test the model's ability to handle severe interference. Blur processing simulates motion blur or focus shifts, ensuring stability in dynamic scenes or low-resolution images.

Random cropping and padding modify the proportions and distributions of objects in images, improving the model's ability to handle occlusion and boundary variations. Random cropping retains the target region while removing peripheral background, simulating scenarios where objects are partially obscured. Padding adds borders to maintain uniform image sizes and introduces background noise, improving boundary detection robustness.

Mosaic data augmentation is an innovative technique introduced in YOLOv5, combining four images into a single composite image to increase sample diversity and inter-object relationships. Mosaic enhances the variability of object positioning and background complexity within images, effectively simulating multi-object scenarios. This method also reduces redundancy between background and objects, providing representative and challenging training samples to improve detection performance in complex industrial environments.

3.4.2. Model Optimization and Deployment

The optimization and deployment of deep learning models are critical to ensuring efficient performance in real-world scenarios. Effective optimization strategies enhance detection accuracy, inference speed, and robustness, while efficient deployment integrates optimized models seamlessly into practical systems, addressing real-time and resource constraints. In this study, model optimization involves techniques such as transfer learning and hyperparameter tuning, while deployment strategies focus on compatibility with embedded devices and industrial applications.

Transfer learning leverages pre-trained models to significantly reduce training costs and accelerate convergence. YOLOv5 provides pre-trained weights based on large-scale datasets (e.g., COCO) as a foundation. By fine-tuning these pre-trained models for automotive component damage detection, the system achieves high detection accuracy with minimal additional data. Key steps include freezing the initial layers of the backbone network to retain generic features and adjusting subsequent layers for task-specific adaptations, enabling rapid performance enhancement under limited datasets.

Hyperparameter tuning optimizes critical parameters during training to enhance model performance. Key parameters such as learning rate, batch size, anchor dimensions, and loss function weights are adjusted. A cosine annealing strategy dynamically adjusts the learning rate, ensuring rapid convergence initially and avoiding local optima later. Anchor dimensions are recalculated based on the actual distribution of damage regions to improve small object detection accuracy. Techniques like grid search and Bayesian optimization automate the search for optimal hyperparameter configurations, maximizing performance across scenarios.

To improve inference speed and reduce memory usage, this study applies model pruning and quantization. Pruning removes network weights and channels with minimal contributions to detection results, reducing model size. Quantization converts model parameters from 32-bit floating-point to 8-bit integer representations, significantly lowering computational costs. TensorRT and similar inference acceleration frameworks further optimize runtime efficiency, meeting real-time requirements on embedded devices.

The optimized YOLOv5 model is deployed on an NVIDIA Jetson Nano platform. Featuring strong edge computing capabilities and support for CUDA acceleration and TensorRT optimization, the Jetson Nano achieves fast inference while conserving system resources. Its compact size and low power consumption make it suitable for industrial and edge device applications, as illustrated in Figure 5.

Fig 5. NVIDIA Jetson Nano Illustration

To ensure deployment efficiency, the optimized YOLOv5 model is converted from PyTorch format to ONNX and subsequently to a TensorRT engine. This conversion process significantly improves inference speed while leveraging hardware acceleration on embedded platforms. The deployment phase includes dynamic input size adaptation to accommodate various component dimensions and scenarios. The core task involves seamlessly integrating the inference module with other system functions, enabling real-time detection and feedback. High-performance APIs invoke the inference engine to return detection results, including bounding boxes and class labels, to the main control system. To address evolving industrial environments, the system supports online model updates and adaptive optimization. By periodically collecting new data for incremental training, the system continually adapts to novel damage types and detection requirements. The online optimization module dynamically adjusts model weights and inference parameters based on field feedback, ensuring consistent detection performance.

Through transfer learning, hyperparameter tuning, and model compression, the system achieves a balance between detection accuracy and inference efficiency. Deployment on the embedded platform significantly enhances real-time capabilities and resource utilization. Testing results indicate that the optimized system accurately detects subtle surface damage in automotive components under complex industrial conditions, fulfilling real-time detection and feedback requirements. This provides a reliable and effective technological solution for industrial manufacturing and quality control.

4. CONCLUSION

This study proposes a modular automotive component damage detection and recognition system that integrates stereo vision and deep learning technologies. The primary achievements include designing a comprehensive system encompassing data acquisition, preprocessing, model training and optimization, and system deployment, as well as achieving efficient detection of subtle damage in complex scenarios through YOLOv5 optimization. The study underscores the potential of stereo vision and deep learning in industrial applications, particularly in 3D reconstruction and object detection tasks. Its practical significance lies in enhancing quality inspection efficiency, supporting real-time detection and dynamic optimization, and providing technological support for industrial automation and intelligent manufacturing. Future research could explore improving model robustness across diverse scenarios and investigate the possibilities of multimodal information fusion and adaptive system optimization. Overall, this research offers valuable insights into the practical application of intelligent detection technologies in the industrial domain, showcasing the promising prospects of combining deep learning and computer vision techniques.

REFERENCES

[1] Roy, Arunabha M., and Jayabrata Bhaduri. "DenseSPH-YOLOv5: An automated damage detection model based on DenseNet and Swin-Transformer prediction head-enabled YOLOv5 with attention mechanism." Advanced Engineering Informatics 56 (2023): 102007.

[2] Zhao, Sizeng, Fei Kang, and Junjie Li. "Concrete dam damage detection and localisation based on YOLOv5s-HSC and photogrammetric 3D reconstruction." Automation in Construction 143 (2022): 104555.

[3] Shang, Yifan, et al. "StereoYOLO: A Stereo Vision-Based Method for Maritime Object Recognition and Localization." Journal of Marine Science and Engineering 12.1 (2024): 197.

[4] Rachidi, Oumayma, Chafik Ed-Dahmani, and Badr Bououlid Idrissi. "A stereo-vision system for real-time person detection in ADAS applications using a fine-tuned version of YOLOv5." Bulletin of Electrical Engineering and Informatics 14.1 (2025): 250-260.

[5] Kustu, Tacettin, and Alev Taskin. "Deep learning and stereo vision based detection of post-earthquake fire geolocation for smart cities within the scope of disaster management: İstanbul case." International journal of disaster risk reduction 96 (2023): 103906.

[6] Xiao, Dong, et al. "A detection method of spangle defects on zinc-coated steel surfaces based on improved YOLO-v5." The International Journal of Advanced Manufacturing Technology 128.1-2 (2023): 937-951.

[7] Zhao, Yue, et al. "GRP-YOLOv5: An improved bearing defect detection algorithm based on YOLOv5." Sensors 23. 17 (2023): 7437.