# Research on Surface Defect Detection Algorithm of Lithium Battery Based on Multi-Modal Deep Learning

**Yuze Yang**
*Jinan Jinqiu International School, Jinan, Shandong, China*

**Abstract: With the rapid advancement of new energy technologies, lithium-ion batteries as core energy storage components have been widely adopted in electric vehicles, portable devices, and energy storage systems, becoming a vital component of modern energy solutions. However, surface defects in lithium-ion batteries significantly compromise their performance, lifespan, and safety-particularly under high-energy-density conditions and frequent usage scenarios where these defects may trigger short circuits, overheating, or even explosions. Current traditional detection methods for lithium-ion battery surface defects primarily rely on manual inspection or conventional computer vision technologies. These approaches demonstrate notable limitations in precision and efficiency, especially when handling various defect types and performing complex defect detection in intricate scenarios. To address the challenges in surface defect detection of lithium batteries, this study proposes an intelligent detection algorithm based on multimodal deep learning. We have designed and implemented an innovative model framework that combines image processing with text analysis to comprehensively improve the accuracy and robustness of defect detection. This method utilizes deep learning models to extract visual features of surface defects and incorporates defect-related textual descriptions as auxiliary information, significantly enhancing the precision in identifying and localizing various types of defects. Experimental results show that the proposed detection algorithm achieves an accuracy of 95.6%, representing a significant performance improvement over existing methods, especially demonstrating strong adaptability and generalization capability in complex scenarios. This research not only provides an efficient and precise technical solution for defect detection in lithium battery production but also offers important theoretical and methodological support for other industrial defect detection tasks.**

**Keywords: Multimodal; Surface Defect Detection; Lithium Battery; Deep Learning**

## 1. The Purpose and Significance of the Topic Selection

### 1.1 Research Background

With the global demand for new energy technologies growing steadily, lithium batteries-as core energy storage components-have been widely adopted in electric vehicles, portable devices, and energy storage systems, emerging as a pivotal element in modern energy solutions. The rapid advancement of electric vehicles and renewable energy technologies in recent years has further fueled the surge in demand for lithium battery applications.[1] Lithium batteries have become the core driving technology for electric vehicles, smart hardware, home appliances and large energy storage systems due to their advantages such as high energy density, long cycle life, light mass and low self-discharge rate[2] However, with the increasing market demand and the continuous expansion of manufacturing scale, quality control problems in lithium battery production have gradually emerged, especially in the production of high energy density batteries, where surface defects are particularly prominent. These defects may have a direct and serious impact on the overall performance and safety of batteries.[3]

There are many kinds of surface defects of lithium battery, including but not limited to scratches, dents, bright spots, decarbonization, bubbles, metal leakage, etc. These defects not only affect the electrochemical performance of the battery (such as capacity, charge and discharge efficiency, cycle life, etc.), but also may cause serious safety risks such as thermal runaway, short circuit or even battery explosion.[4] The emergence of surface defects on batteries, particularly in critical applications like electric vehicles and energy storage systems, poses serious risks to user safety, equipment reliability, and environmental protection. In high-density and high-energy battery systems, localized temperature spikes caused by surface defects may trigger thermal runaway, potentially leading to severe safety incidents.[5-7] Therefore, ensuring the integrity of the battery surface, timely detection and accurate positioning of these defects is one of the core issues to ensure the safe operation of batteries.

In recent years, with the continuous optimization of manufacturing processes, surface defect detection in lithium batteries has gradually become an important part of lithium battery production. Traditional defect detection methods mainly rely on manual visual inspection and automated detection technologies based on traditional image processing.[8] While manual inspection can achieve initial defect detection, its reliance on human expertise results in low efficiency, poor accuracy, and difficulty meeting the high-efficiency demands of mass production. Traditional image processing-based defect detection methods, though effective in specific scenarios, often depend on image quality, lighting conditions, and predefined feature extraction rules. This makes them prone to limitations when dealing with various complex environments and diverse defects.[9-10]

In view of the limitations of traditional detection methods, deep learning technology, especially convolutional neural network (CNN), has made remarkable breakthroughs in the field of image processing in recent years[11] Deep learning methods can automatically learn efficient and abstract features from a large amount of data, overcome the difficulties of manual feature extraction, and have strong adaptability and generalization ability[12] However, single-image data still falls short in handling complex defects, particularly when dealing with subtle, blurry, or indistinguishable flaws against the background. In such cases, image information alone may lack sufficient diagnostic evidence. Consequently, multimodal deep learning technology has emerged as a crucial solution in recent years. By integrating multiple information sources like images and text, this approach significantly enhances detection accuracy and robustness.[13]

In the task of surface defect detection of lithium battery, the combination of image mode and text mode can effectively make up for the deficiency of single mode[14] Image data can provide visual information of surface defects, while text data contains descriptive information such as defect type, size and location. Especially in describing complex or small defects, text information can provide valuable contextual support for the model[15] Through the multimodal deep learning model, it can comprehensively utilize the detailed information in the image and the semantic information in the text to achieve more accurate and efficient defect detection and localization in different defect types and complex production environments.

The research on surface defect detection of lithium battery based on multi-modal deep learning can effectively improve the automation level of the defect detection system, reduce the need for manual intervention, reduce production costs, and greatly improve the accuracy and consistency of detection[16] While ensuring battery quality, this approach also provides robust technical safeguards for lithium battery safety. As the lithium battery market continues to expand and technology advances, deep learning-based multimodal defect detection technology is increasingly becoming an indispensable core component in lithium battery production processes.

## 1.2 Research Status of Defect Detection

### 1.2.1 Traditional defect detection

In the field of defect detection, traditional methods and modern machine learning techniques each have their own advantages and are suitable for different application scenarios. Traditional image processing methods rely on inherent image features such as edges, textures, shapes, and gray-level variations to achieve defect detection. For example, Jia et al.[17] An automatic fabric defect detection method based on lattice segmentation and template statistics is proposed, which achieves a detection rate of up to 97% by comparing the similarity between the lattice and statistical templates. Kumari et al[18] A defect detection method based on Sylvester matrix similarity estimation is developed. The test image is aligned with the reference image through geometric transformation and image resampling, and the detection is carried out by comparing the rank of Sylvester matrix. Qiu et al[19] A defect detection method suitable for non-uniform illumination is proposed, which combines significance detection with inherent image decomposition to effectively improve the detection accuracy of low-contrast defects. Zhang et al[20] The proposed CrackNet defect detection method, which consists of four layers of convolutional layers with the same width and height dimensions but different channel numbers, outputs pixel-level defect detection results; Yu et al[21] Two full convolutional FCN networks are used for defect detection. The first FCN network is responsible for coarse inference of defect location, and the second FCN network is responsible for refinement of defect detection results. The algorithm is verified in the defect data set.

With the development of technology, machine learning methods have been gradually introduced into defect detection, especially the support vector machine (SVM) method. Due to its powerful feature learning and classification ability, it has become one of the research hotspots. For example, Song et al[22] The MCITF algorithm is proposed to combine the improved texture features with Laplacian regularization technology. Through superpixel segmentation and saliency map generation, background interference is effectively reduced and the accuracy of surface defect detection on steel strip is significantly improved. Zhou et al[23] The WR-IFOA-SVM method is proposed, which combines u-LBP and GLCM texture features, and further improves the accuracy of wire rope surface defect detection by optimizing SVM parameters.

1.2.2 Defect detection based on deep learning

The current research trend is to combine deep learning with traditional machine learning to compensate for the limitations of a single approach. Singh et al[24] By integrating a pre-trained ResNet-101 convolutional neural network (CNN) with multiple support vector machines (SVMs), this hybrid approach effectively detects defects in non-grinding processes through data augmentation and network layer activation analysis. This method not only reduces dependence on extensive training datasets but also significantly enhances detection efficiency and accuracy. Deep learning-based defect detection has become one of the most widely adopted technologies in lithium battery surface inspection, primarily categorized into three types: region candidate-based target detection methods, regression-based target detection methods, and search-based target detection methods [25,26]. Region candidate-based detection methods include R-CNN[27], FastR-CNN[28], FasterR-CNN[29] and MaskR-CNN[30] These methods achieve target detection and localization by extracting candidate regions and then classifying and bordering each candidate region. Lu et al[31] A real-time defect detection and closed-loop adjustment method based on deep learning is proposed, using FasterR-CNN, SSD and YOLOv4 models to detect fiber path misalignment and wear in the printing process in real time. The average accuracy (mAP) of the three algorithms is 73.28%,65.29% and 90.42% respectively. Zhang et al[32] An improved Faster R-CNN method is proposed for defect detection in 3D printed lattice structures. An efficient defect detection model is constructed by K-medoids algorithm, adaptive anchor selection based on Manhattan distance, data enhancement and fine-tuning strategies. Experimental results show that the average accuracy of the model reaches 93.4% in defect detection.

The improved MaskR-CNN algorithm has made significant progress in the field of defect detection. Xia et al[33] An improved surface defect detection method has been proposed by integrating a CBAM attention module into the Region Proposal Network (RPN) and combining it with a Path Aggregation Feature Pyramid Network (FPN), effectively fusing multi-level features. The method achieves 90% accuracy on defect datasets, demonstrating superior detection performance compared to traditional approaches. Xu et al.[34] Surface defect detection is carried out based on the enhanced MaskR-CNN, especially by introducing a path-enhanced feature pyramid network and edge detection branch into the network. Compared with the traditional computer vision method based on manual features, the detection accuracy and effect are significantly improved. Wang et al[35] Then, a new multi-scale fusion feature pyramid structure is designed to optimize the MaskR-CNN network, and the traditional IoU is replaced by the full intersection and union ratio (IoU). In the experiment, the average accuracy is 98.70%, which further improves the accuracy of defect detection.

As an end-to-end regression target detection method, YOLO series model has been widely used in defect detection tasks[36] This method can detect targets efficiently by dividing input images into grid cells and returning the existence, position and category of targets in each cell. Yuan et al[37] A novel PCB surface defect detection method (YOLO-HMC) based on an enhanced version of YOLOv5 has been proposed. By integrating the HorNet architecture and MCBAM to boost feature extraction and defect localization capabilities, combined with CARAFE technology to optimize the upsampling layer, the system significantly enhances image information aggregation. Through optimized detection head design, the model parameters are reduced, achieving an average precision of 98.6% in PCB defect detection. Zhang et al.[38] Based on YOLOv3, the K-medoids clustering algorithm and super-resolution convolutional neural network were used to significantly improve the recognition ability of small defects, and the average accuracy of 94.55% was achieved in the final experiment. Qian et al[39] By integrating ShuffleNetv2 into the feature extractor and introducing a lightweight feature pyramid network, the parameter quantity of YOLO model was successfully reduced while improving the efficiency of multi-scale feature fusion, achieving an average accuracy of 79.23% on NEU-DET dataset. Wen et al[40] The YOLOv7 model has been enhanced by incorporating the Convolution Block Attention Module (CBAM) and Adaptive Spatial Feature Fusion (ASFF) techniques. These improvements enable the model to better capture image details and fuse features across multiple scales. When tested on the FCC and BCC-FCC datasets, the accuracy rate reached 96.9%, representing a 2.4% improvement over the original model. Jiang et al.[41] The K-Means++ algorithm is used to optimize the anchor box, which effectively alleviates the problem of imbalance between defect aspect ratio, and a new attention module combining maximum pooling and average pooling is proposed to improve the detection accuracy of small defects.

In defect detection, supervised learning methods often demonstrate limited effectiveness, while unsupervised and weakly supervised learning approaches show tremendous potential. However, in industrial applications, the high cost of obtaining labeled data typically makes it difficult to acquire sufficient defect samples for training.[42] In order to overcome this challenge, transfer learning has been widely studied as an effective solution. Badmos et al[43] By comparing the defect detection methods of transfer learning and non-transfer learning, the experiment shows that the transfer learning based detection method achieves 99% F1 index, significantly better than the method without transfer learning. Damacharla et al[44] When training with the U-net framework, we compared defect detection performance between pre-trained and untrained models using ImageNet. Results demonstrated that the transfer learning-based model achieved 26% higher accuracy on our self-built dataset compared to non-transfer learning counterparts. These findings highlight the effectiveness of transfer learning in defect detection, particularly valuable for industrial applications where annotated data is scarce.

### 1.2.3 Defect detection with multimodal feature fusion

In recent years, with the rapid development of deep learning and computer vision technologies, multimodal learning has gained increasing attention in defect detection applications. The core concept of multimodal feature fusion lies in effectively integrating information from various modalities (such as images, text, sensor data, etc.) to enhance models' capability in detecting complex defects.[45] Traditional single-modal defect detection methods typically rely on analyzing only one type of feature, which limits their performance in complex scenarios. In contrast, multimodal feature fusion effectively combines the strengths of multiple information sources, thereby overcoming the limitations of single-modal data.[46-48] Recent studies have shown that combining images with other types of data (such as sound, temperature sensor data, etc.) can significantly improve the accuracy and robustness of defect detection.

Research on multimodal feature fusion has made significant progress in many fields. Lu et al[49] A defect detection method combining multimodal fusion convolutional neural networks with cross-attention mechanisms has been proposed. Through feature extraction, fusion, and defect analysis modules, the accuracy of defect detection is significantly improved. Asad et al. [50] developed the 2M3DF network that integrates multi-view RGB images and point cloud information, effectively enhancing 3D industrial defect detection performance. Experimental results demonstrate that this network achieves outstanding detection capabilities while providing real-time outputs. Cheng et al. [51] introduced BatteryGPT-a multimodal large language model (MLLM)-based lithium-ion battery defect detection system. By integrating image encoders, text encoders, and a large language model, the method detects anomalies in battery images using minimal normal samples. Experimental results show BatteryGPT achieves an AUC value of 96.49% in battery defect detection, surpassing existing models CutPaste and SCADN, thus offering a more efficient solution for battery inspection. Feng et al. [52] proposed a novel image-text multimodal fusion deep learning network to improve insulator defect detection accuracy. This approach combines text and image data through a transformer framework with six attention mechanisms, integrates feature pyramid networks for multi-scale image feature extraction, and enhances color features via feature differentiation and principal component analysis to capture details and contextual information. Simultaneously, the proposed method employs the WordPiece approach for text processing. Experimental results demonstrate that the proposed method significantly enhances accuracy and robustness in complex environments, with recall rates 14% to 31.7% higher than traditional methods. Zhai et al. [53] developed a novel One-Stage MFFNet for pavement crack detection, which integrates boundary box coordinates and mask information to improve detection accuracy and segmentation integrity. Experimental results on self-collected datasets and public datasets (CFD and CRACK500) show that MFFNet achieves 2.6% higher detection accuracy and 4.7% better segmentation accuracy compared to Mask R-CNN, while outperforming the optimized RDSNet model by 1.8% in detection accuracy and 2FPS in processing speed. The experimental results indicate that MFFNet demonstrates optimal performance in both detection and segmentation accuracy, making it a high-precision pavement crack detection model.

### 1.3 The Main Research Content of This Subject

The primary objective of this research is to develop a lithium battery surface defect detection algorithm based on multimodal deep learning. By integrating image and text information, the algorithm aims to achieve efficient and accurate identification and localization of various surface defects in lithium batteries. This system can detect and locate typical defects including scratches, depressions, bright spots, decarburization, bubbles, and metal leakage, meeting the stringent requirements for defect detection in lithium battery manufacturing. Additionally, the project explores methods to enhance model robustness under limited or zero samples, enabling stronger generalization capabilities to handle previously unseen defect types and achieve rapid response to anomalies. The research on lithium battery surface defect detection algorithm based on multimodal deep learning consists of four parts. First, the construction of a lithium battery surface defect dataset. Second, the development of multimodal feature fusion methods. Third, the design and optimization of deep learning models. Fourth, model evaluation and application verification. These parts are carried out in sequence, forming a complete research process for lithium battery surface defect detection.

The main work of this paper is divided into the following three aspects:

Construction of lithium battery surface defect data set: Collect and collate the image data set of lithium battery surface containing various defect types. The data is collected by industrial CCD camera and annotated with text description to ensure the comprehensiveness and diversity of the data set;

Research on Multimodal Feature Fusion Methods: A key focus of this study is the design and optimization of multimodal feature fusion approaches. Since images and text data originate from different modalities and each contain distinct semantic information, relying on a single modality proves insufficient for complex defect detection. To address this challenge, the research will explore multimodal feature interaction mechanisms and develop an effective cross-modal fusion method. This approach will organically integrate image features with text features, achieving efficient fusion of both visual and textual information through deep neural networks.

Design and Optimization of Deep Learning Models: This research focuses on developing and refining deep learning models for lithium battery surface defect detection. By leveraging advanced architectures such as CNN and Transformers, we explore how to enhance detection accuracy and robustness through integrated modeling of image and text features. Furthermore, cutting-edge feature extraction techniques including Feature Pyramid Networks (FPN) are employed to improve the models capability in detecting defects of varying sizes and morphological configurations.

Model Evaluation and Application Validation: After completing the design and training of the algorithmic model, the project will conduct systematic evaluation and performance verification. The assessment will be performed through testing on both self-built datasets and public datasets (such as MVTec-AD and VisA), comparing the performance of different models in defect detection. Key metrics including detection accuracy, recall rate, and F1-score will be evaluated, with comparisons made against existing mainstream defect detection methods. Ultimately, the project will validate the practical effectiveness and application value of the proposed method in lithium battery production through simulated industrial scenarios.

## 2. Research Program

### 2.1 Proposed Technical Solutions

This study proposes a multi-modal deep learning-based framework for lithium battery surface defect detection, aiming to enhance accuracy, localization precision, and generalization capabilities. The framework comprises three core modules: SPCI module, DAEP module, and CMI module. First, the SPCI module captures row-level and column-level features of image block tokens to precisely locate defects, demonstrating strong adaptability to directional defects like linear and crack-like patterns commonly found on lithium battery surfaces. By enhancing spatial structure modeling capabilities, this module improves model performance in complex defect scenarios. Next, the P-M module integrates global visual features with text cues through dynamic attention mechanisms, reducing dependence on specific categories and boosting adaptability to diverse anomaly types. Particularly effective for detecting rare or unknown defects, this module enhances practicality and scalability in real-world applications. Finally, the CMI module achieves deep interaction between image and text features by capturing local details and global context while optimizing text embedding, significantly improving cross-modal understanding and information fusion capabilities. This integrated approach effectively elevates overall performance in complex. The system achieves stable anomaly detection in dynamic defect images. Furthermore, the architecture is built on large-scale pre-trained visual language models (e.g., CLIP), utilizing image encoders and text-guided mechanisms to enable unsupervised or zero-shot defect detection. This approach not only applies to lithium battery surface defect detection but also handles multimodal data in similar industrial inspection scenarios. Experimental verification and comparative analysis will be conducted on both public datasets and a self-built lithium battery image dataset to evaluate the proposed models accuracy, recall rate, and robustness in practical applications. The research aims to provide an efficient and precise solution for defect detection in lithium battery production processes. As illustrated in Figure 1, this constitutes the methodology of our study.

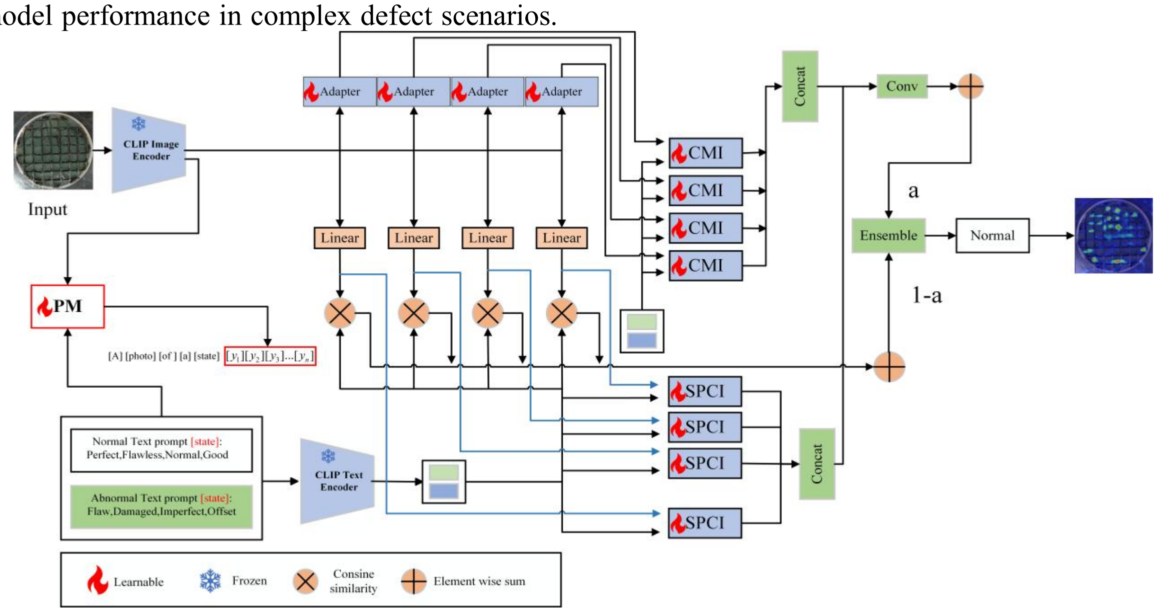


**Figure 1. Lithium Battery Surface Defect Detection Algorithm based on Multimodal Deep Learning**

 

### 2.2 Conditions Required for the Implementation of the Plan

2.2.1 Hardware condition

The hardware requirements for this project primarily involve configuring high-performance computing resources and data acquisition equipment. To ensure efficient training and inference processes for deep learning models, it is essential to equip them with high-performance GPUs such as NVIDIA Tesla or RTX series graphics cards, which accelerate model training and enable large-scale data processing. To meet computational demands, servers or workstations with robust computing capabilities are recommended.

2.2.2 Software environment

This research will utilize deep learning frameworks such as TensorFlow or PyTorch to develop and optimize multimodal deep learning models, which support the integration of complex Convolutional Neural Networks (CNNs) with Natural Language Processing (NLP) systems. Image data preprocessing and augmentation serve as critical training steps, employing tools like OpenCV or PIL for denoising, cropping, and rotation to enhance dataset diversity and model robustness. Textual data will be processed through NLP tools such as spaCy or Transformers to extract semantic features from defect descriptions, thereby improving the models ability to comprehend textual information.

2.2.3 Experimental data set

The success of this research hinges on high-quality datasets. The lithium battery surface defect dataset was self-collected using industrial CCD cameras for precise imaging. This comprehensive collection includes various defects such as scratches, dents, bright spots, decarburization, bubbles, and metal leaks-all representing common surface flaws encountered in lithium battery manufacturing. To ensure data diversity and representativeness, every defect type in the dataset mirrors actual production scenarios, covering abnormalities of varying sizes, shapes, and severity levels. This approach provides robust training data for developing subsequent defect detection algorithms.

### 2.3 Main Problems and Solutions

2.3.1 The difficulty of integrating image and text data

In multimodal deep learning, effectively integrating image and text data remains a major challenge. While images provide visual information, text descriptions capture defect characteristics and locations. The key challenge lies in combining both approaches to enhance detection accuracy and robustness, particularly when dealing with complex defects or significant background interference.

Solution: To address this challenge, this paper proposes a multimodal fusion framework that utilizes Convolutional Neural Networks (CNN) for image feature extraction and Natural Language Processing (NLP) for text information extraction. The framework employs a deep attention mechanism to dynamically balance the contributions of image and text modalities, while employing multi-level feature fusion to enhance recognition capabilities for complex backgrounds and defects.

2.3.2 Robustness of defect detection in complex environments

The surface defect detection of lithium battery may be affected by uneven illumination, complex background and small defects, resulting in poor robustness of traditional methods. Especially in the production environment, equipment and conditions change greatly, so the detection system needs strong adaptability to ensure stability and accuracy.

Solution: To enhance robustness, this study implements a data augmentation strategy that simulates diverse lighting conditions, background variations, and defect scenarios to improve model adaptability. By integrating CNN and RNN, the approach leverages both image and temporal data features to strengthen environmental adaptation capabilities. Furthermore, self-supervised learning is employed to boost generalization performance when training with limited labeled datasets.

2.3.3 Lack of defect labeling data

The defects in the production process of lithium battery are complex and varied, resulting in insufficient high-quality annotated data, which affects the training and generalization ability of deep learning model.

Solution: To address the data scarcity issue, this paper employs data augmentation techniques to generate composite defect images and expand the dataset. We further integrate transfer learning by fine-tuning pre-trained models from other domains. Additionally, a self-supervised learning strategy is implemented where the model generates its own labels, significantly reducing reliance on manual annotations.

## 3. Expected Research Results

### 3.1 Expected Project Objectives and Results

The research objective of this project is to develop an intelligent detection algorithm for lithium battery surface defects based on multimodal deep learning, and apply it to an automated defect detection system in lithium battery manufacturing. This method integrates image and text data, leveraging deep learning models to enhance the accuracy and robustness of surface defect detection. The aim is to achieve efficient, automated defect detection, reduce manual intervention, and improve production efficiency and quality. Specific objectives include: improving detection accuracy through multimodal data fusion methods, striving to increase precision by 20% across various types and complex backgrounds; while optimizing model computational efficiency to enable real-time or near-real-time detection processes in industrial production, significantly accelerating inspection speed.

Academic achievements:

Two SCI or EI papers: "Research on Surface Defect Detection Algorithm of Lithium Battery Based on Multimodal Deep Learning" and "Application of Multimodal Deep Learning in Industrial Defect Detection: Taking Surface Defect Detection of lithium battery as an example".

Two related patents: "A method for surface defect detection of lithium battery based on multimodal deep learning" and "A deep learning algorithm for surface defect localization of lithium battery".

Soft copyright: 1. "A lithium battery surface defect detection software".

### 3.2 Practical Problems to be Solved

This study introduces a multimodal deep learning approach that integrates image data with text descriptions to develop an efficient and accurate defect detection algorithm. The algorithm automatically identifies and locates various surface defects. By optimizing the models adaptability to diverse defects, it enhances automation levels and detection accuracy while reducing manual intervention. The research not only addresses limitations of traditional detection methods but also provides intelligent, automated solutions for lithium battery production quality control. This advancement improves manufacturing efficiency and product quality, supporting technological innovation and industrial applications in the battery industry.

## 4. Conclusions and Future Work

This research develops a multimodal deep learning algorithm to address surface defect detection in lithium battery production. By combining image and text data, the study proposes an innovative framework that integrates visual and semantic features, significantly improving detection accuracy and robustness. The system can accurately identify defects like scratches and dents, maintaining high performance in complex environments, thus driving automation and quality control in the production process.

Unlike traditional methods, which rely on manual inspection or conventional image processing, this approach leverages CNN and transformers to achieve precise detection, even for minor defects. The use of data augmentation, transfer learning, and self-supervised learning enhances the model's generalization ability in environments with limited data. The results provide an efficient solution for lithium battery defect detection and offer valuable insights for similar industrial applications, advancing smart manufacturing and quality control in the new energy sector.

## References

[1] Honghong Deng. Research on Surface Defect Detection of Lithium Batteries [D]. Harbin Institute of Technology, 2018.

[2] Hanwen Yu and Yiquan Wu . Research Progress on Machine Vision-Based Defect Detection of Lithium Batteries [J]. Journal of Instrumentation, 2024,45(9):1-23.

[3] Xueyong Feng. A Deep Learning-Based Method for Surface Defect Recognition and Classification of cylindrical Lithium Battery Steel Shells [D]. Hefei University of Technology, 2020.

[4] Liu X, Wu L, Guo X, Liu X, Wu L, Guo X, et al. A novel approach for surface defect detection of lithium battery based on improved K-nearest neighbor and Euclidean clustering segmentation[J]. The International Journal of Advanced Manufacturing Technology, 2023, 127(1): 971-985.

[5] Xu C, Li L, Li J, Xu C, Li L, Li J, et al.

 

Surface defects detection and identification of lithium battery pole piece based on multi-feature fusion and PSO-SVM[J]. Ieee Access, 2021, 9: 85232-85239.
[6] Chen X, Liu M, Niu Y, Chen X, Liu M, Niu Y, et al. Deep-learning-based lithium battery defect detection via cross-domain generalization[J]. IEEE Access, 2024.
[7] Chen W, Han X, Pan Y, et al. Defects in lithium-ion batteries: From origins to safety risks[J]. Green Energy and Intelligent Transportation, 2025, 4(3): 100235.
[8] Huang P, Wang C, Yang S, et al. Review of Research on Battery Defect Detection and Recovery[C]//International Workshop of Advanced Manufacturing and Automation. Singapore: Springer Nature Singapore, 2024: 148-155.
[9] Jia Z, Wang M, Zhao S. A review of deep learning-based approaches for defect detection in smart manufacturing[J]. Journal of Optics, 2024, 53(2): 1345-1351.
[10] Jiahui Chen, Fei Wang, et al. Research progress on nondestructive testing technology of lithium battery safety performance[J]. Nondestructive Testing, 44(12): 72-75.
[11] Tang B, Chen L, Sun W, et al. Review of surface defect detection of steel products based on machine vision[J]. IET Image Processing, 2023, 17(2): 303-322.
[12] Chen Y, Shu Y, Li X, et al. Research on detection algorithm of lithium battery surface defects based on embedded machine vision[J]. Journal of Intelligent & Fuzzy Systems, 2021, 41(3): 4327-4335.
[13] Rahmati M, Rahmati N. A multimodal deep learning framework for real-time defect recognition in industrial components using visual, acoustic and vibration signals[J]. Journal of Intelligent Man ufacturing and Special Equipment, 2025: 1-20.
[14] Cheng D, Wang S, Li C, et al. BatteryGPT: Battery Anomaly Detection Based on Multimodal Large Language Model[C]//2024 5th International Conference on Power Engineering (ICPE). IEEE, 2024: 516-521.
[15] Jeong J, Zou Y, Kim T, et al. Winclip: Zero-/few-shot anomaly classification and segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 19606-19616.
[16] Ning Y, Yang F, Zhang Y, et al. Bridging multimodal data and battery science with machine learning[J]. Matter, 2024, 7(6): 2011-2032.
[17] Jia L, Chen C, Xu S, et al. Fabric defect inspection based on lattice segmentation and template statistics[J]. Information Sciences, 2020, 512: 964-984.
[18] Kumari R, Bandara G, Dissanayake M B. Sylvester Matrix-Based Similarity Estimation Method for Automation of Defect Detection in Textile Fabrics[J]. Journal of Sensors, 2021, 2021(1): 6625421.
[19] Qiu Y, Tang L, Li B, et al. Uneven illumination surface defects inspection based on saliency detection and intrinsic image decomposition[J]. IEEE Access, 2020, 8: 190663-190676.
[20] Zhang A, Wang K C P, Li B, et al. Automated pixel-level pavement crack detection on 3D asphalt surfaces using a deep-learning network[J]. Computer-Aided Civil and Infrastructure Engineering, 2017, 32(10): 805-819.
[21] Yu Z, Wu X, Gu X. Fully convolutional networks for surface defect inspection in industrial environment[C]//Computer Vision Systems: 11th International Conference, ICVS 2017, Shenzhen, China, July 10-13, 2017, Revised Selected Papers 11. Springer International Publishing, 2017: 417-426.
[22] Song G, Song K, Yan Y. Saliency detection for strip steel surface defects using multiple constraints and improved texture features[J]. Optics and Lasers in Engineering, 2020, 128: 106000.
[23] Zhou P, Zhou G, Li Y, et al. A hybrid data-driven method for wire rope surface defect detection[J]. IEEE Sensors Journal, 2020, 20(15): 8297-8306.
[24] Singh S A, Desai K A. Automated surface defect detection framework using machine vision and convolutional neural networks[J]. Journal of Intelligent Manufacturing, 2023, 34(4): 1995-2011.
[25] Gui S, Song S, Qin R, et al. Remote sensing object detection in the deep learning era-a review[J]. Remote Sensing, 2024, 16(2): 327.
[26] Manakitsa N, Maraslidis G S, Moysis L, et al. A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision[J].

Technologies, 2024, 12(2): 15.
[27] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.
[28] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
[29] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.
[30] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
[31] Lu L, Hou J, Yuan S, et al. Deep learning-assisted real-time defect detection and closed-loop adjustment for additive manufacturing of continuous fiber-reinforced polymer composites[J]. Robotics and Computer-Integrated Manufacturing, 2023, 79: 102431.
[32] Zhang Y, Zhang Z, Fu K, et al. Adaptive defect detection for 3-D printed lattice structures based on improved faster R-CNN[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-9.
[33] Xia B, Luo H, Shi S. Improved Faster R-CNN Based Surface Defect Detection Algorithm for Plates[J]. Computational intelligence and neuroscience, 2022, 2022(1): 3248722.
[34] Xu Y, Li D, Xie Q, et al. Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN[J]. Measurement, 2021, 178: 109316.
[35] Wang H, Li M, Wan Z. Rail surface defect detection based on improved Mask R-CNN[J]. Computers and Electrical Engineering, 2022, 102: 108269.
[36] Ali M L, Zhang Z. The YOLO framework: A comprehensive review of evolution, applications, and benchmarks in object detection[J]. Computers, 2024, 13(12): 336.
[37] Yuan M, Zhou Y, Ren X, et al. YOLO-HMC: An improved method for PCB surface defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-11.
[38] Zhang Z, Zhang Y, Wen Y, et al. Intelligent defect detection method for additive manufactured lattice structures based on a modified YOLOv3 model [J]. Journal of Nondestructive Evaluation, 2022, 41: 1-14.
[39] Qian X, Wang X, Yang S, et al. LFF-YOLO: A YOLO algorithm with lightweight feature fusion network for multi-scale defect detection [J]. IEEE Access, 2022, 10: 130339-130349.
[40] Wen Y, Cheng J, Ren Y, et al. Complex defects detection of 3-D-Printed lattice structures: Accuracy and scale improvement in YOLO V7[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-9.
[41] Jiang L, Yuan B, Wang Y, et al. MA-YOLO: a method for detecting surface defects of aluminum profiles with attention guidance[J]. IEEE Access, 2023, 11: 71269-71286.
[42] Lei H, Wang B, Wu H, et al. Defect Detection for Polymeric Polarizer Based on Faster R-CNN[J]. J. Inf. Hiding Multim. Signal Process., 2018, 9(6): 1414-1420.
[43] Badmos O, Kopp A, Bernthaler T, et al. Image-based defect detection in lithium-ion battery electrode using convolutional neural networks[J]. Journal of Intelligent Manufacturing, 2020, 31: 885-897.
[44] Damacharla P, Rao A, Ringenberg J, et al. TLU-net: a deep learning approach for automatic steel surface defect detection[C]//2021 International Conference on Applied Artificial Intelligence (ICAPAI). IEEE, 2021: 1-6.
[45] Chazhoor A A P, Hu S, Gao B, et al. MDC-Net: Multimodal Detection and Captioning Network for Steel Surface Defects[C]//International Conference on Robotics, Computer Vision and Intelligent Systems. Cham: Springer Nature Switzerland, 2024: 316-333.
[46]Wang R, Du W, Jiang Q. Quantitative estimation method for complex part surface defects based on multimodal information fusion[J]. Complex & Intelligent Systems, 2025, 11(6): 1-27.
[47] Zhao Z, Hu B, Feng Y, et al. Multi-surface defect detection for universal joint bearings via multimodal feature and deep transfer learning[J]. International Journal of Production Research, 2023, 61(13): 4402-4418.

 

[48] Wang H, Peng R, Zhu Y, et al. Few-Shot Photovoltaic Film Defect Detection With Contextual Ensemble Language-Image Multimodal Network[J]. IEEE Transactions on Industrial Informatics, 2025.
[49] Lu H, Zhu Y, Yin M, et al. Multimodal fusion convolutional neural network with cross-attention mechanism for internal defect detection of magnetic tile[J]. IEEE Access, 2024, 10: 60876-60886.
[50] Asad M, Azeem W, Jiang H, et al. 2M3DF: Advancing 3D Industrial Defect Detection with Multi Perspective Multimodal Fusion Network[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2025.
[51] Cheng D, Wang S, Li C, et al. BatteryGPT: Battery Anomaly Detection Based on Multimodal Large Language Model[C]//2024 5th International Conference on Power Engineering (ICPE). IEEE, 2024: 516-521.
[52] Feng B, Xia X, Zhang L, et al. An image-text multimodal fusion of deep learning for detecting insulator defects[J]. International Journal of Parallel, Emergent and Distributed Systems, 2025: 1-17.
[53] Zhai J, Sun Z, Huyan J, et al. Automatic pavement crack detection using multimodal features fusion deep neural network[J]. International Journal of Pavement Engineering, 2023, 24(2): 2086692.