

A Review of the Application of the ResNet18 Model in Medical and Industrial Visual Inspection

Zehou Liu

*School of Automation, Central South University, Changsha, China
8207221825@csu.edu.cn*

Abstract. This article systematically reviews the research status, key technological breakthroughs, and application prospects of the ResNet18 model and its improved variants in the fields of medical imaging and industrial vision detection. ResNet18 effectively alleviates the deep network degradation problem with its residual structure, providing an efficient benchmark model for tasks such as image classification, object detection, and defect recognition. This article focuses on analyzing the improvement strategies of ResNet18 for different application scenarios, such as introducing attention mechanisms (SE, SimAM), combining transfer learning, data augmentation, and model lightweight pruning, which significantly enhance performance in tasks including coronary artery narrowing classification, road pothole detection, metal defect evaluation, flame stability recognition, epilepsy EEG classification, and coal gangue sorting, achieving high accuracy levels generally ranging from 94% to 99.5%. Additionally, this article discusses challenges such as model lightweight deployment, few-shot learning, cross-domain generalization, and standardized dataset construction, and looks forward to future research directions, including deeper structural optimization, multimodal fusion, and embedded applications in broader industrial scenarios, providing comprehensive technical references for researchers in related fields.

Keywords: ResNet18, deep learning, image recognition, attention mechanism, transfer learning, lightweight network

1. Introduction

With the rapid advancement of artificial intelligence technology, deep learning has demonstrated tremendous potential in the fields of image recognition and intelligent inspection, becoming a core driving force behind the transformation of medical diagnosis and industrial automation. Among various deep networks, ResNet18, with its introduced residual connection structure, effectively addresses the issues of gradient vanishing and degradation in deep network training. It combines moderate model depth, high computational efficiency, and superior performance, making it a favored benchmark model for many visual tasks. In the medical field, its precise analysis of medical images can assist doctors in improving diagnostic efficiency and accuracy. In industrial scenarios, vision inspection systems based on ResNet18 significantly enhance the automation level and reliability of defect recognition and quality control. Therefore, systematically reviewing the

application status of ResNet18 and its improved models across various domains, as well as summarizing their optimization strategies and performance, holds significant theoretical and engineering value for promoting the practical implementation of artificial intelligence technology in real-world scenarios.

In recent years, researchers have made various improvements and optimizations to the ResNet18 model to meet the specific requirements of different application scenarios, achieving remarkable results. For example, Ye et al. [1] combined ResNet18 with transfer learning for the binary classification task of coronary artery stenosis images. This approach directly utilized a pre-trained model for fine-tuning, achieving a final classification accuracy of 96.97% without complex structural adjustments, significantly outperforming the baseline model. This demonstrates the powerful transfer capability and effectiveness of ResNet18 in medical image processing. In industrial inspection, Ma et al. [2] addressed the specific challenges of classifying laser images of road potholes by introducing a channel attention mechanism (Squeeze-and-Excitation, SE) into ResNet18, constructing an SE-ResNet18 model. This mechanism adaptively calibrates channel feature responses, enhancing focus on critical targets. The model ultimately achieved an accuracy of 99.58% and an F1-score of 93.78% in the pothole classification task, highlighting the advantages of attention mechanisms in improving feature representation. Furthermore, to meet the demands of real-time industrial inspection, Xue et al. [3] explored lightweight deployment of the model. They used ResNet18 as the backbone feature extraction network for the YOLOv3 detection algorithm and incorporated unstructured pruning for model compression. This approach achieved a mAP of 96.27% and a single-frame detection speed of 45.5 milliseconds in coal gangue sorting tasks, significantly improving detection efficiency while maintaining high accuracy. This provides a successful example for deploying algorithms on resource-constrained embedded devices.

These studies fully demonstrate the powerful adaptability and potential for enhancement of the ResNet18 model when combined with strategies such as transfer learning, attention mechanisms, and model pruning. This paper aims to provide a systematic review of the application progress of ResNet18 and its improved models in the medical and industrial fields, analyzing the technical characteristics and applicable scenarios of different improvement strategies, with the goal of offering valuable references for research and practice in related domains.

2. Chapter 2 overview of related technologies

As an important member of the deep residual network family, ResNet18 has become one of the benchmark models in the fields of medical image analysis and industrial visual inspection due to its excellent feature extraction capability and good computational efficiency. This chapter will systematically review various improved models based on ResNet18 and their application results across different fields in recent years, focusing on analyzing their improvement strategies, technical characteristics, and performance.

2.1. Application of ResNet18 model based on transfer learning

Transfer learning can effectively address the issue of scarce labeled data in specific domains, significantly enhancing the model's generalization ability on small datasets. Ye et al. applied this strategy to the field of medical imaging and explored the performance of ResNet18 in binary classification tasks of coronary artery stenosis images. This method directly fine-tunes the ResNet18 model pre-trained on large datasets such as ImageNet to adapt to the feature distribution of medical images. Experimental results show that through transfer learning alone, the classification accuracy

of the proposed method reaches 96.97%, which is an improvement of 8.08% compared to a ResNet18 model trained from scratch, demonstrating the immense potential of transferring pre-trained models in medical image diagnosis and providing an efficient and reliable solution for medical auxiliary diagnosis with limited samples.

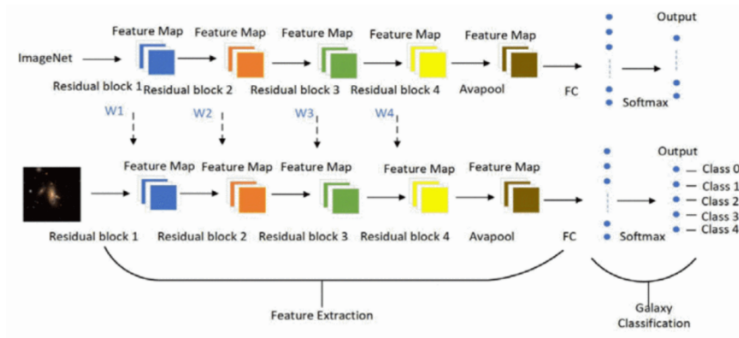


Figure 1. ResNet18 network model based on transfer learning [4]

2.2. Improved model with integrated attention mechanism

To enhance the model's ability to focus on key features, the attention mechanism has been widely introduced to improve the original ResNet18 architecture.

2.2.1. Se-ResNet18 model

To address the challenges of small pothole targets and complex backgrounds in laser road images, Ma et al. embedded a channel attention module (Squeeze-and-Excitation, SE) into ResNet18, constructing an SE-ResNet18 model. This module explicitly models inter-channel dependencies to adaptively recalibrate channel-wise feature responses, enabling the model to allocate more computational resources to feature channels rich in information. Ablation studies confirmed that this mechanism effectively enhances the model's perception of pothole features. The approach ultimately achieved an accuracy of 99.58% and an F1-score of 93.78% in the classification task, significantly outperforming other mainstream image classification models. It provides an effective technical solution for fine-grained target classification in complex backgrounds [2].

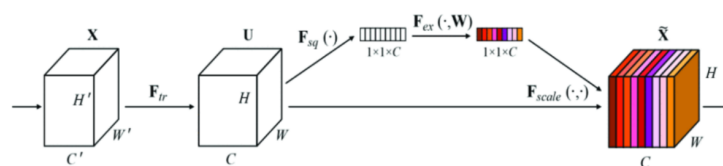


Figure 2. Channel attention module network module [5]

2.2.2. ResNet18-SimAM modle

In the task of boiler flame stability detection in power plants, Wang et al. integrated a parameter-free SimAM (Simple Parameter-Free Attention Module) attention mechanism into the lower layers of ResNet18. The SimAM module assigns weights to neurons without requiring additional parameters, effectively enhancing the expression of critical features while suppressing redundant information. Experimental results demonstrated that the improved ResNet18-SimAM model not only achieved faster convergence but also significantly improved accuracy, reaching 97.54% on the training set and 94.01% on the test set. This provides a high-precision detection tool for condition monitoring and safety early warning in industrial scenarios [3].

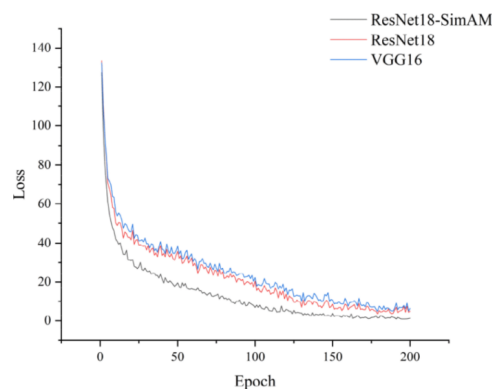


Figure 3. Loss curves of ResNet18-SimAM, ResNet18, and VGG16 during training [6]

2.3. A lightweight improved model for real-time detection

In industrial deployment, detection speed and model size are often critical constraints. To address the stringent real-time requirements of coal gangue sorting robots, Xue et al. proposed a lightweight detection model based on ResNet18. The study utilized ResNet18 as the backbone feature extraction network for the YOLOv3 object detection algorithm, replacing the original Darknet-53, which significantly reduced the number of model parameters. To further compress the model, the researchers introduced unstructured pruning technology, eliminating redundant connections in the network and achieving extreme model compression. Ultimately, this lightweight model maintained high precision (with mAP reaching 96.27%) while achieving a detection speed of 45.5 milliseconds per frame and a model size of only 65.34 MB. Compared to the original YOLOv3 algorithm, the detection rate improved by 39.3%, and the number of parameters was reduced by 71.4%, fully meeting the industrial requirements for real-time performance and embedded deployment [7].

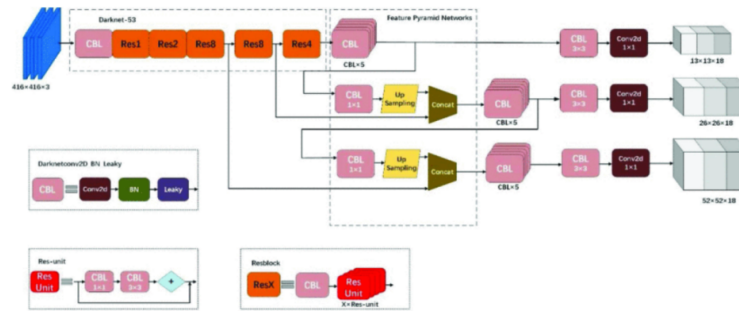


Figure 4. YOLOv3 network structure [8]

2.4. Application of special data preprocessing techniques

In certain specialized domains, raw data may not consist of natural images and must be transformed to be effectively processed by convolutional neural networks. In a study on classifying epileptic seizure phases based on electroencephalogram (EEG) signals, Dubey et al. innovatively adopted a data preprocessing strategy. The research first segmented and reassembled the raw EEG signals to perform data augmentation, then employed continuous wavelet transform (CWT) to convert the one-dimensional time-series signals into two-dimensional time-frequency spectrograms, thereby transforming the signal classification problem into an image classification task. The converted spectrograms were processed using a standard ResNet-18 model, ultimately achieving accuracies of 98.4%, 99.1%, and 98.2% in three classification tasks (ictal/interictal, normal/ictal, and normal/interictal, respectively). This demonstrates ResNet18's strong capability in handling multimodal data when combined with specific preprocessing techniques [9].

In summary, by integrating transfer learning, attention mechanisms, model pruning, and specific data preprocessing strategies, ResNet18 continues to adapt to the complex and diverse application demands of the medical and industrial fields. It achieves an excellent balance among accuracy, speed, and model size, demonstrating remarkable scalability and practical utility.

Table 1. Comparison of applications of various models of ResNet18 model

Serial number	Model name	Improvement method	Application field	Accuracy/performance indicators
1	ResNet18 Transfer Learning	Join transfer learning	Classification of coronary artery stenosis images	Accuracy 96.97%
2	SE-ResNet18	Introduce channel attention mechanism	Laser image classification of road surface potholes	Accuracy 99.58%, F1 score 93.78%
3	Improved ResNet18	The convolutional layer is replaced by three convolutions with a kernel size of two.	Assessment of Defect Depth in Metal Plates	Aluminum plate defect depth accuracy 98.1%, stainless steel 95.4%.
4	ResNet18-SimAM	Integrate SimAM attention mechanism	Stability Testing of Power Plant Boiler Flames	Training set 97.54%, test set 94.01%
5	ResNet18	Data augmentation Continuous wave transformation	Classification of Epileptic Seizure Stages (EEG)	Three sets classification accuracy: 98.4%, 99.1%, 98.2%
6	ResNet18-YOLO	Unstructured pruning YOLOv3 improvement	Identification and selection of coal gangue	mAP 96.27%, detection time 45.5ms, model size 65.34MB

3. Introduction to application scenarios

ResNet18 and its improved models, with their strong feature extraction capabilities and high computational efficiency, have been successfully deployed in numerous practical scenarios in medical and industrial fields, providing effective technical solutions for specific issues in particular industries.

3.1. Medical imaging assisted diagnosis

In the medical field, the ResNet18 model is widely used for assisted diagnosis across various types of medical imaging, aiming to improve diagnostic accuracy and efficiency while reducing the workload of physicians. In a study by Ye et al., the model was applied to cardiovascular disease diagnosis. By fine-tuning a pre-trained ResNet18 model, automatic classification of coronary angiographic images was achieved, accurately determining the presence of vascular stenosis with an accuracy of 96.97% [1]. This technology can serve as an auxiliary tool for clinicians in large-scale preliminary screenings, facilitating early detection of lesions and providing patients with valuable treatment time. Additionally, Dubey et al. innovatively applied ResNet18 to neurological disease diagnosis [3]. Their study utilized continuous wavelet transform to convert electroencephalogram (EEG) signals into time-frequency spectrograms, which were then processed by ResNet18 to achieve high-precision classification of epileptic states—ictal, interictal, and normal—with the highest accuracy reaching 99.1% in one of the three classification tasks. This provides a new method for objective and automated diagnosis of epilepsy, holding significant clinical value.

3.2. Industrial vision and automated inspection

In the industrial sector, ResNet18 and its variants act as "intelligent quality inspectors" on production lines, enabling smart perception of product defects and equipment conditions.

3.2.1. Infrastructure maintenance

Ma et al. applied an SE-ResNet18 model integrated with a channel attention mechanism to road condition monitoring [2]. This model accurately classifies road surface images obtained via laser scanning to automatically identify defects such as potholes, achieving an accuracy of 99.58%. It provides data support for preventive road maintenance, enabling timely detection of potential hazards, ensuring traffic safety, and reducing maintenance costs.

3.2.2. Smart manufacturing and non-destructive testing

She et al. combined an improved ResNet18 model with a four-coil array eddy current sensor to evaluate the depth of defects in metal sheets [3]. This solution can quantitatively assess defect depths in aluminum and stainless steel plates with high precision (achieving accuracies of 98.1% and 95.4%, respectively), serving as an advanced non-contact non-destructive testing method. It is particularly valuable for industries with stringent material quality requirements, such as aerospace and automotive manufacturing.

3.2.3. Energy production safety monitoring

Wang et al. utilized a ResNet18-SimAM model to monitor the stability of flame states in power plant boiler furnaces [6]. By analyzing flame images, the model can determine combustion stability in real time, achieving an accuracy of 94.01% on the test set. This application plays a critical role in ensuring the safe and efficient operation of power plants, preventing combustion accidents, and contributing to energy conservation and emissions reduction.

3.2.4. Resource sorting and industrial robotics

Xue et al. focused on applying ResNet18 in the field of resource recycling [7]. They developed a lightweight ResNet18-YOLO-based algorithm for coal gangue sorting, integrated into sorting robots. This system can identify and separate gangue from coal in real time, achieving a mean average precision (mAP) of 96.27% with a processing time of only 45.5 milliseconds per frame. This not significantly improves sorting efficiency and accuracy while reducing labor costs but also promotes the intelligent and clean development of coal mining.

4. Challenge analysis

4.1. Contradiction between model efficiency and real-time requirements [7]

Challenge: Many industrial scenarios (e.g., sorting robots, online quality inspection) impose stringent millisecond-level response time requirements. However, high-precision models often come with complex structures and a large number of parameters, leading to high computational overhead and slow inference speeds. This makes real-time deployment on embedded devices or edge computing platforms with limited computational resources challenging.

Solution: Model compression techniques (e.g., pruning [6], quantization, knowledge distillation) can be employed to reduce parameter counts and computational demands. Designing more efficient network architectures (e.g., using depthwise separable convolutions) and combining hardware acceleration can further enhance inference speed, striking an optimal balance between accuracy and efficiency.

4.2. Domain data scarcity and class imbalance issues [1, 9]

Challenge: In the medical field and certain industrial niches (e.g., specific types of defects), acquiring large volumes of high-quality, annotated data is difficult and costly. Additionally, the phenomenon where normal samples vastly outnumber defective samples is extremely common. Such severe class imbalance can cause models to overly favor the majority class, making it difficult to effectively learn the features of the minority class (defects), thereby compromising the model's generalization ability and detection accuracy.

Solution: Data augmentation techniques (e.g., rotating, scaling, adding noise to images) can be used to expand the training set. For class imbalance, strategies such as resampling (oversampling the minority class or undersampling the majority class), introducing class weights into the loss function, or using generative adversarial networks (GANs) to synthesize realistic minority class samples can help mitigate the issue.

4.3. Insufficient model generalization and adaptive capability [2, 3]

Challenge: A model trained on a specific dataset and environment may experience significant performance degradation when applied to a slightly different scenario (e.g., changes in lighting conditions, equipment models, or material types). For instance, a defect detection model trained on one type of metal plate may perform poorly on another. This lack of generalization limits the model's universality and large-scale adoption.

Solution: Using more diverse datasets during training, employing domain adaptation techniques, and leveraging transfer learning for continuous fine-tuning can help the model better adapt to new target scenarios, enhancing its robustness and generalization performance.

4.4. Computational resource and deployment cost constraints [6, 7]

Challenge: Improved models (e.g., those incorporating attention mechanisms) often increase computational complexity, demanding higher hardware processing power and memory from deployment devices. For many small and medium-sized enterprises, this translates to higher hardware procurement and maintenance costs, creating barriers to practical implementation.

Solution: On one hand, further algorithmic optimization can be pursued to achieve "lightweight" improvements. On the other hand, exploring cloud and edge collaborative deployment schemes—where complex model inference tasks are handled in the cloud while the edge focuses on data collection and simple processing—can help optimize cost structures.

References

- [1] X. Ye, Y. Jin, Z. Wang, N. Feng, L. Mu and G. Xie, "Image Binary Classification of Coronary Artery Stenosis Based on Resnet18 and Transfer Learning, " in 2023 35th Chinese Control and Decision Conference (CCDC), Yichang, China, 2023, pp. 2721-2726, doi: 10.1109/CCDC58219.2023.10326590.
- [2] R. Ma, M. Xu, Z. Li and X. Sun, "Classification of Pothole Laser Images Based on the Improved ResNet18 Network, " in 2024 3rd International Conference on Electronics and Information Technology (EIT), Chengdu,

China, 2024, pp. 619-623, doi: 10.1109/EIT63098.2024.10762034.

- [3] S. She et al., "Evaluation of Defects Depth for Metal Sheets Using Four-Coil Excitation Array Eddy Current Sensor and Improved ResNet18 Network, " in IEEE Sensors Journal*, vol. 24, no. 12, pp. 18955-18967, 15 June15, 2024, doi: 10.1109/JSEN.2024.3367816.
- [4] Y. Guo, P. Liu, X. Mei and Z. Yang, "The Research on Galaxy Image Classification Model Based on ResNet18 and Transfer Learning, " 2024 8th International Workshop on Control Engineering and Advanced Algorithms (IWCEAA), Nanjing, China, 2024, pp. 139-142, doi: 10.1109/IWCEAA63616.2024.10823737.
- [5] N. Shahadat, "Lung Image Analysis using Squeeze and Excitation Based Convolutional Networks, " 2023 26th International Conference on Computer and Information Technology (ICCIT), Cox's Bazar, Bangladesh, 2023, pp. 1-6, doi: 10.1109/ICCIT60459.2023.10441501.
- [6] Z. Wang, Y. Bian, M. Yang and G. Liu, "Power Plant Furnace Flame Stability Detection Based on ResNet18-SimAM, " in 2024 IEEE Sustainable Power and Energy Conference (iSPEC), Kuching, Sarawak, Malaysia, 2024, pp. 321-325, doi: 10.1109/iSPEC59716.2024.10892397.
- [7] G. Xue, S. Li, P. Hou, S. Gao and R. Tan, "Study on a Resnet18-Based Lightweight Recognition ALgorithm of Coal and Gangue, " in 2024 International Conference on Identification, Information and Knowledge in the Internet of Things (IIKI), Kusatsu, Japan, 2024, pp. 83-87, doi: 10.1109/IIKI65561.2024.00024.
- [8] Y. Yu, S. Yan and X. Hao, "Pedestrian Detection Based on Improved YOLOv3 Network, " 2023 IEEE International Conference on Control, Electronics and Computer Technology (ICCECT), Jilin, China, 2023, pp. 297-301, doi: 10.1109/ICCECT57938.2023.10141270. keywords: {Roads; Feature extraction; Real-time systems; Periodic structures; Pedestrian detection; YOLOv3; Convolutional Block Attention Module; SPP Structure},
- [9] V. K. Dubey, S. Sarkar, R. Shukla, G. Singh and A. Sahani, "Epileptic Seizure Stage Classification from EEG Signal Using ResNet18 Model and Data Augmentation, " in 2022 IEEE Delhi Section Conference (DELCON), New Delhi, India, 2022, pp. 1-5, doi: 10.1109/DELCON54057.2022.9753352.