Steel Surface Defect Detection Using Faster R-CNN with Traditional Image Feature Extraction Methods

Yongzhuo Yang
Fudan University
Shanghai, China
22307140078@m.fudan.edu.cn

Ruifeng Chen
Fudan University
Shanghai, China
22307140061@m.fudan.edu.cn

Weiqi Xu
Fudan University
Shanghai, China
21302010059xx@gmail.com

Abstract

Steel detection plays a critical role in ensuring the quality and safety of steel products, which are widely used across various industries. This study investigates an enhanced approach to steel defect detection using the Faster R-CNN (Region-based Convolutional Neural Network) architecture. The traditional Faster R-CNN model is modified by incorporating a dual-channel backbone, where the second channel includes four distinct image enhancement techniques: CLAHE (Contrast Limited Adaptive Histogram Equalization), LBP (Local Binary Pattern), FFT (Fast Fourier Transform), and Sobel filtering. These enhancements aim to provide additional feature representations that may improve the detection of subtle defects in steel surfaces. Additionally, the loss function is modified through the use of focal loss, addressing class imbalance. The potential impact on the overall detection performance of these modifications is shown through experiments.

1. Introduction

1.1. Background

The quality of steel production is crucial for industrial applications, as defects on steel surfaces can compromise structural integrity, safety, and customer satisfaction. Automated detection and classification of surface defects are essential to ensure high-quality standards. Recent advancements in computer vision, particularly deep learning, have enhanced the detection and classification of various defects, including rolled-in scale (RS), patches (Pa), cracks (Cr), pitted surface (PS), inclusions (In), and scratches (Sc).

The task of steel defect detection is particularly challenging due to several factors. A primary difficulty lies in the reliance on deep learning methods, such as ResNet-50, for feature extraction. While these models are powerful, they may overlook critical features, especially subtle defects, as they learn high-level representations that do not always capture low-level, yet important, details. This can result in the model missing key information that is crucial for accurate defect detection. Additionally, there is significant variation in the difficulty of detecting different defect categories. Some defects, such as cracks and inclusions, are subtle and difficult to distinguish from the surrounding surface, while others, like patches and scratches, may appear more obvious but still exhibit variations in size, shape, and texture. Moreover, the limited availability of labeled training data exacerbates the issue, as deep learning models typically require large and diverse datasets to generalize effectively. These challenges highlight the need for more robust methods that can address both the feature extraction limitations and the data scarcity, improving the model's ability to detect diverse and subtle defects.

1.2. Solution

To address the challenges in steel defect detection, we propose three key strategies aimed at improving model performance:

Dual-Channel Backbone: In the Faster R-CNN architecture, we use ResNet-50 as the backbone for feature extraction. However, deep learning models, while effective, may overlook certain critical features due to the complexity of learned representations. To overcome this limitation, we modify the backbone structure by adding a second channel. In this additional channel, we employ four traditional image processing methods—Sobel filtering, Local Binary Pattern (LBP), Fast Fourier Transform (FFT), and Contrast Limited Adaptive Histogram Equalization (CLAHE)—to extract complementary features. These manually engineered features are then fused with the deep learning-based features in the main backbone, aiming to enhance the model's ability to detect various defects and improve overall performance.

Data Augmentation for Difficult Classes: To enhance the detection of harder-to-identify defect categories, we apply data augmentation techniques, such as rotation, flipping, and scaling, to artificially increase the diversity and quantity of samples for these specific classes. This approach aims to improve the accuracy of the model in detecting defects that are subtle or visually similar to the background.

Modified Loss Function: We modify the traditional loss function by implementing focal loss, which helps to focus more on the difficult-to-classify samples and reduces the model's sensitivity to easy-to-classify ones. This adjustment is especially important for addressing class imbalance, ensuring that the model gives appropriate attention to underrepresented and harder-to-detect defect categories.

These strategies are designed to address the key challenges of class difficulty and limited data while enhancing the model's ability to capture both learned and hand-crafted features for more robust defect detection.

1.3. Key Contributions

Key contributions of this study include:
- Data augmentation techniques targeting underrepresented classes.
- Focal loss adjustments to emphasize challenging defect categories.
- A dual-channel backbone integrating four traditional image processing techniques: FFT, Sobel filtering, LBP, and CLAHE.

2. Related Work

2.1. Traditional Methods

Early defect detection methods relied on manual inspection and basic image processing. Machine learning models, such as support vector machines (SVM) and decision trees, were later employed but were limited by handcrafted features and inability to handle complex variations in defect appearance [3].

2.2. Deep Learning Approaches

Deep learning has advanced defect detection through automated feature extraction and robust classification. Faster R-CNN, YOLO, and RetinaNet have demonstrated high accuracy in object detection tasks. Faster R-CNN, with its region proposal network (RPN), is particularly effective for precise localization and classification [4].

2.3. Class Imbalance

Class imbalance is a persistent challenge, as some defect types occur less frequently in datasets. Techniques like focal loss and data augmentation address this issue by reweighting class contributions and expanding underrepresented classes [2].

2.4. Dual-Channel Backbone

In this study, a dual-channel backbone integrates both spatial and frequency-domain features to enhance steel defect detection. The Dual-Channel Backbone employs the following four traditional image enhancement techniques:
- **Sobel Filtering:** Highlights gradient changes to extract edge features, aiding in delineating defect boundaries.
- **Local Binary Pattern (LBP):** Analyzes micro-texture variations, enhancing sensitivity to subtle surface irregularities .
- **Fast Fourier Transform (FFT):** Transforms images into the frequency domain, identifying periodic patterns and sensitivity variations critical for defect detection [3].
- **Contrast Limited Adaptive Histogram Equalization (CLAHE):** Enhances local contrast, particularly in regions with uneven illumination, to highlight defect characteristics.

By fusing these enhanced features with deep learning-based spatial features in a dual-channel architecture, the model achieves a comprehensive representation of defect characteristics. This approach addresses the limitations of traditional methods and supports robust detection across diverse defect types [1].

3. Method

3.1. Data Augmentation

To address the issue of class imbalance, we applied targeted data augmentation techniques. Images with underrepresented defect classes were duplicated and horizontally flipped to enhance the representation of challenging defect types. This approach ensures a more balanced dataset and improves model generalization.

3.2. Focal Loss Design

Focal loss was employed to mitigate the impact of class imbalance further. By introducing a dynamic scaling factor, the loss function emphasizes harder-to-classify samples while reducing the influence of easily classified ones. The focal loss function is defined as:

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t), \tag{1}$$

where p_t is the predicted probability for the correct class, α_t is a weighting factor for class imbalance, and γ adjusts the rate at which easy examples are down-weighted. This formulation, as introduced in [2], enhances the model's ability to detect underrepresented and challenging defect types.

3.3. Dual-Channel Backbone

A dual-channel backbone was developed to integrate both spatial-domain and enhanced feature representations for improved defect detection. The first channel processes the original input images, allowing the model to extract standard spatial features using ResNet-50. The second channel, on the other hand, processes the images with four traditional image enhancement techniques: Sobel filtering, LBP,

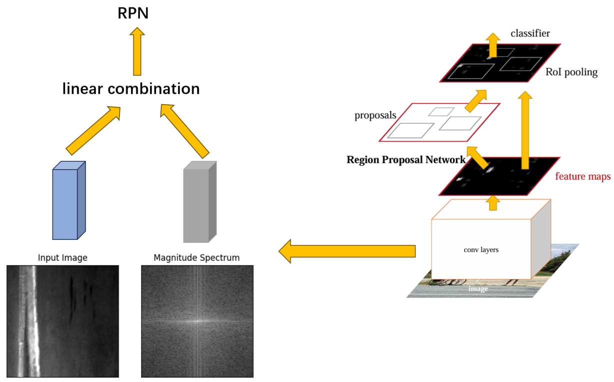

Figure 1. Architecture of the dual-channel backbone, showcasing the integration of spatial and frequency-domain features.

FFT, and CLAHE. These methods aim to highlight different characteristics of the defects, such as edges, textures, and contrast variations. The features from both channels are then fused, potentially providing a more comprehensive representation of defect characteristics. We hypothesize that this dual-channel approach may help the model better capture subtle and diverse defects, improving detection performance.

4. Experiment

4.1. Data Augmentation

To address the challenge of limited training data and improve the model's performance on underrepresented classes, we applied targeted data augmentation techniques. Specifically, when the proportion of a difficult-to-detect class in a given image exceeded half, we created a duplicate of the image and applied a horizontal flip. This strategy aims to enhance the model's ability to recognize challenging defect categories by artificially increasing their presence in the training dataset, helping the model learn to identify these defects more effectively.

As shown in Figure 2 and Figure 3, the mAP on the validation set before data augmentation was slightly higher than after augmentation. However, in the test set, the model with augmented data outperformed the one without augmentation. This indicates that while data augmentation initially causes a slight decrease in validation performance, it significantly improves the model's generalization ability, leading to better performance on unseen data.

Method	mAP (%)	Loss
Without Data Augmentation	34.99	1.630
With Data Augmentation	35.22	1.151

Table 1. Comparison of performance before and after data augmentation.

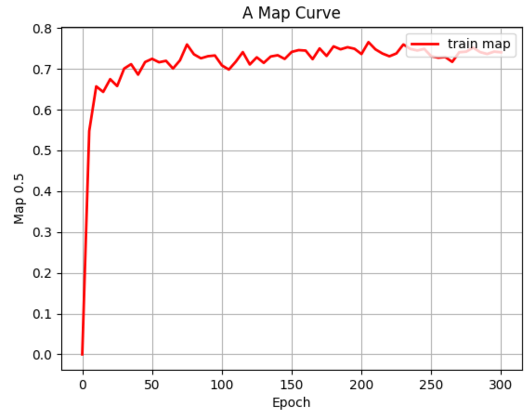

Figure 2. mAP on Validation Set Before Data Augmentation

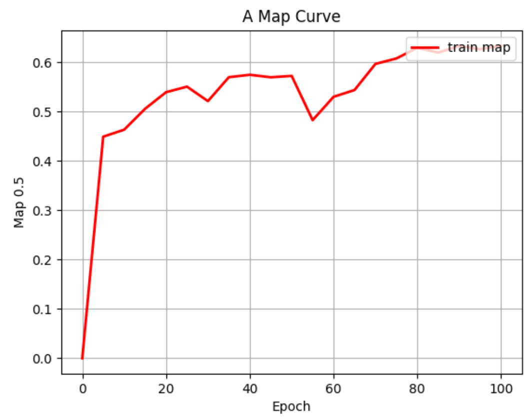

Figure 3. mAP on Validation Set After Data Augmentation

4.2. Focal Loss

In this study, we used focal loss to address the class imbalance problem, particularly for defects that are difficult to detect. Focal loss modifies the standard cross-entropy loss function by down-weighting easy-to-classify samples and focusing more on the hard examples. We conducted experiments to compare the performance of the model with and without focal loss. Unfortunately, this method did not succeed, and using focal loss resulted in a decrease in performance.

Method	mAP (%)	Loss
Without Focal Loss	34.99	1.630
With Focal Loss	34.37	1.071

Table 2. Comparison of performance before and after applying focal loss.

4.3. Image Fusion

In our experiments, we first tested the performance of the model with direct image fusion as a preliminary step before exploring the dual-channel approach. In this experiment, we fused each of the processed images (generated using one of the four enhancement techniques: Sobel, LBP, FFT, and CLAHE) with the original image. The fusion was performed by concatenating the processed image with the original image before passing them through the ResNet-50 backbone. This approach was an attempt to investigate whether direct fusion could successfully improve detection performance. However, results from these experiments showed that while image fusion provided some complementary features, it did not yield substantial improvements in performance.

Method	mAP(%)	Loss
Origin	35.22	1.151
Sobel	30.33	1.441
LBP	27.23	1.377
FFT	33.92	1.287
CLAHE	31.27	1.278

Table 3. Comparison of performance (mAP and Loss) with and without image fusion for different enhancement methods.

4.4. Dual-Channel Backbone

To further enhance defect detection, we proposed a dual-channel backbone. The first channel processes the original input images using ResNet-50 for feature extraction, while the second channel processes the images with traditional feature enhancement techniques, such as Sobel filtering, LBP, FFT, and CLAHE. The outputs of both channels are then fused. Unlike direct fusion, which applies enhanced features directly to the model, the dual-channel architecture allows the model to learn whether these manually engineered features are useful. While the dual-channel approach yielded better results compared to direct fusion, it did not outperform the baseline (Origin), indicating that the model might not have fully benefited from the additional features.

Method	mAP (%)	Loss
baseline	35.22	1.151
dual-Sobel	28.72	1.412
dual-LBP	30.21	1.443
dual-FFT	29.99	1.343
dual-CLAHE	30.02	1.459

Table 4. Comparison of performance (mAP and Loss) between dual-channel backbone and baseline methods.

5. Analysis

5.1. Analysis of Focal Loss

Despite the theoretical benefits of focal loss in addressing class imbalance, our experiments show that it did not improve performance and, in fact, led to a decrease in accuracy. Focal loss is designed to focus more on hard-to-classify examples by down-weighting easy examples, which is expected to improve model performance, especially on imbalanced datasets. However, in our case, the introduction of focal loss increased the difficulty for the model to learn from the majority class and may have caused an overemphasis on subtle defects, which are inherently difficult to detect. This imbalance in learning might have led to poorer generalization, especially for defects that are easier to identify. Moreover, the lack of significant improvement in performance suggests that the model was already capable of handling class imbalance to some extent, rendering the additional emphasis on harder examples unnecessary. As a result, focal loss might have caused instability in training, leading to a higher loss and lower mAP.

5.2. Analysis of Dual-Channel Backbone

The dual-channel backbone, which incorporated both learned features from ResNet-50 and manually engineered features (Sobel, LBP, FFT, CLAHE), did not outperform the baseline model (ResNet-50 only). While the idea was to allow the model to learn whether additional features would be useful, it seems that the additional complexity introduced by the dual-channel approach may have hindered performance. One possible reason for this is that the manually engineered features did not provide complementary information but rather introduced noise or redundancy, which the model could not effectively use. Additionally, the model might have struggled to properly integrate the two channels and learn the relationships between the learned and manually engineered features. In some cases, adding extra channels can complicate the optimization process, leading to suboptimal performance. Therefore, while the dual-channel approach theoretically offered a richer feature set, it did not lead to better performance than the baseline, indicating that more careful feature selection or integration strategies might be necessary to fully exploit the potential of dual-channel architectures.

6. Conclusion

We proposed a Faster R-CNN-based framework for detecting six common types of steel surface defects. To address key challenges such as class imbalance and variability in defect characteristics, we incorporated targeted data augmentation, focal loss adjustments, and a dual-channel backbone with feature fusion. The experimental results showed that data augmentation effectively improved the model's performance, particularly for underrepresented defect classes. However, the focal loss adjustments and dual-channel feature fusion did not yield the expected improvements. These findings suggest that while data augmentation enhances the model's ability to generalize, further refinement of loss functions and feature fusion techniques is needed. Future work will focus on optimizing these components, exploring real-time applications, and improving the model's performance for industrial deployment.

7. Addition

7.1. Author Contributions

Yongzhuo Yang: Completed the baseline, designed the dual-channel code, conducted model training, created the presentation (PPT), and wrote the paper.
Ruifeng Chen: Completed the baseline, performed data augmentation, designed focal loss, implemented feature extraction code, and co-wrote the paper.
Weiqi Xu: Conducted literature review and contributed to the paper writing.

7.2. Final Group Ranking

排名	参赛团队	所属组织	score
1	茫然的小七	复旦大学	43.73005
2	子若非瑜的团队	复旦大学	42.55127
3	戚风蛋糕派的团队	复旦大学	42.48193
4	YongzhouYang的团队	无	35.21758

Figure 4. Final group ranking.

The group achieved a final ranking of 4th out of 6 teams.

7.3. Code Repository

All the code and results at the GitHub link: https://github.com/yatao-zhuozhuo/2024_CV_team20

References

[1] Zhang Q. Li, H. and L. Wang. Frequency perception network for camouflaged object detection. *arXiv preprint arXiv:2308.08924*, 2023. 2

[2] Goyal P. Girshick R. He K. Lin, T.-Y. and P. Dollar. Focal loss for dense object detection. *arXiv preprint arXiv:1708.02002*, 2017. 2

[3] Sridharan N. V. Sugumaran V. Vasan, V. and R. J. Balasundaram. Hot rolled steel surface defect detection and classification using an automatic ensemble approach. *Engineering Research Express*, 6(025544), 2024. 2

[4] Wei X. Chen Z. Zhong, Y. and J. Zhang. Detecting camouflaged object in frequency domain. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2