IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

INTEGRATING MULTI SCALE FEATURE EXTRACTION FOR ROBUST SMALL FOREIGN OBJECT DETECTION USING RESNET AND FASTER R-CNN

[1]Dr. Ragavamsi Davuluri, [2]Konatham Nandini, [3]Chintala Sravya, [4]Gudiseva Denny Siva Sai Siddhu, [5]Machineedi Prem Kumar

[1]Associate Professor, [2]Student, [3]Student, [4]Student, [5]Student
[1]Department of Computer Science and Engineering,
[1]Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, India

Abstract: Accurate detection of small foreign objects is a critical requirement in industrial inspection systems, where minor defects can significantly impact product quality, safety, and operational efficiency. Conventional inspection methods often fail to identify small, low contrast, or partially occluded foreign objects in complex industrial environments. To address these limitations, this work proposes a deep learning based detection framework that integrates multi scale feature extraction with region based object detection. The proposed approach employs a Residual Network as the backbone for hierarchical feature extraction, enhanced by a Feature Pyramid Network to capture discriminative features across multiple scales. Faster R-CNN is utilized for precise localization and classification of foreign objects. Experimental evaluation using precision, recall, F1 score, and mean average precision demonstrates that the proposed framework provides a robust and reliable solution for real time industrial quality inspection applications.

Index Terms **-** Foreign object detection, small object detection, ResNet, Feature Pyramid Network, Faster R-CNN, industrial inspection

I. INTRODUCTION

The detection of foreign objects is a critical requirement in modern industrial inspection systems, where safety, product quality, and operational reliability must be strictly maintained. Industries such as manufacturing, food processing, pharma- ceuticals, and aerospace demand highly accurate inspection mechanisms, as even minor contamination can compromise product functionality and result in safety hazards, financial losses, and reputational damage.

Although industrial automation has advanced significantly, manual visual inspection is still practiced in many envi- ronments. However, such inspection methods are limited by human fatigue, subjectivity, and perceptual constraints, leading to missed detections particularly when defects are small, partially occluded, or visually similar to the background. With the increasing speed of automated production lines, these limitations highlight the need for intelligent and automated detection solutions.

Recent progress in deep learning has significantly improved object detection performance. Convolutional Neural Networks can automatically learn hierarchical feature representations from raw images, enabling robust detection of small, irregular, and low contrast foreign objects in cluttered environments. Among CNN based architectures, Residual Networks have proven effective for feature extraction due to their residual learning mechanism, which allows deep networks to be trained efficiently while preserving fine grained spatial details.

To further address scale variation, Feature Pyramid Net- works are integrated with ResNet to construct multi scale feature representations. For precise localization and classifica- tion, the Faster R-CNN framework is employed. The proposed work presents an intelligent inspection framework designed for challenging industrial environments.

II. LITERATURE REVIEW

Foreign object detection is one area that has received a great deal of attention in industrial inspection and quality control, and it is primarily because of the impact it has on the reliability of the products. The initial research in this area was carried out using conventional image processing. Although computationally efficient, these approaches lack robustness under varying illumination, surface conditions, and object scales.

To overcome these limitations, multi scale feature extraction techniques were introduced. With the advancement of deep learning, CNN based detectors have demonstrated superior performance. Residual Networks enable deeper architectures through residual learning, while Feature Pyramid Networks improve sensitivity to small objects by fusing features from multiple layers.

Faster R-CNN is widely adopted for accurate object lo- calization and classification. Its Region Proposal Network generates candidate regions that are refined through classi- fication and bounding box regression. Existing studies show that integrating ResNet, FPN, and Faster R-CNN provides a robust solution for small foreign object detection in complex environments.

The primary objectives of this research are:

- To design an automated system for detecting small for- eign objects in industrial images.
- To integrate ResNet with Feature Pyramid Networks for effective multi scale feature extraction.
- To employ Faster R-CNN for precise localization and classification.
- To enhance detection robustness under challenging con- ditions such as illumination variation and background complexity.
- To evaluate performance using standard detection metrics.

III. METHODOLOGY

A. Dataset Preparation

The dataset is collected and annotated using the Roboflow platform in COCO format. Defect categories include crazing, inclusion, patches, and scratches. Data augmentation tech- niques such as rotation, flipping, scaling, and illumination adjustment are applied to improve robustness.

B. Image Preprocessing

The images are resized to a certain scale. Augmentation techniques resemble the real-world industrial environment. This helps in generalization.

C. Proposed System

The proposed system is based on the following approaches:

ResNet for feature extraction,Feature Pyramid Network for feature representation,Faster R-CNN for detection

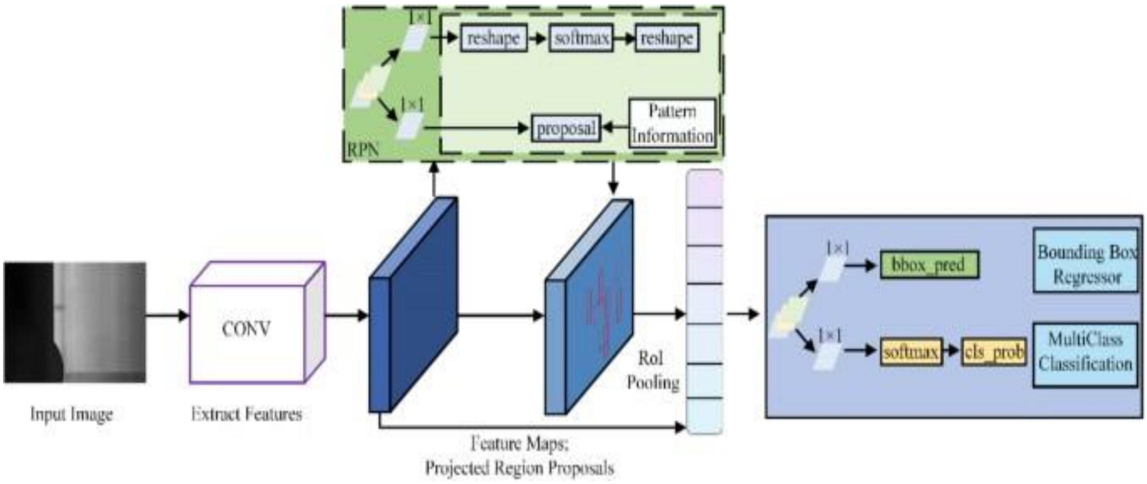

Fig. 1. Proposed Small Foreign Object Detection System

D. System Architecture

The proposed system architecture is designed as a modular and scalable multi-model detection framework that integrates deep learning based object detection and anomaly detection techniques. The architecture consists of a user interface, backend inference controller, trained model repository, parallel detection models, a fusion engine, and a visualization module. The workflow begins at the User Interface, where users upload industrial images and view detection results. The up- loaded image is forwarded to the Backend API and Inference Controller, which manages request handling, input validation, and inference coordination. This controller acts as the central processing unit of the system.

The backend retrieves pretrained models from the Trained Model Repository. The repository stores multiple trained mod- els, enabling parallel execution without redundant loading. Three complementary models are employed: Faster R-CNN with ResNet and Feature Pyramid Network for precise local- ization, YOLOv8 for fast and efficient object detection, and an Autoencoder-based anomaly detection model for identifying unseen or irregular defects.

Each model processes the input image independently and produces detection outputs in the form of bounding boxes, confidence scores, and heatmaps. These outputs are aggregated by the Fusion Engine, which applies Non-Maximum Suppres- sion and confidence-based weighting to eliminate redundant detections and improve reliability.

The fused output provides the final defect type and its spatial location, which is returned to the User Interface for visualiza- tion. This architecture enhances detection accuracy, robustness, and scalability for industrial inspection applications.

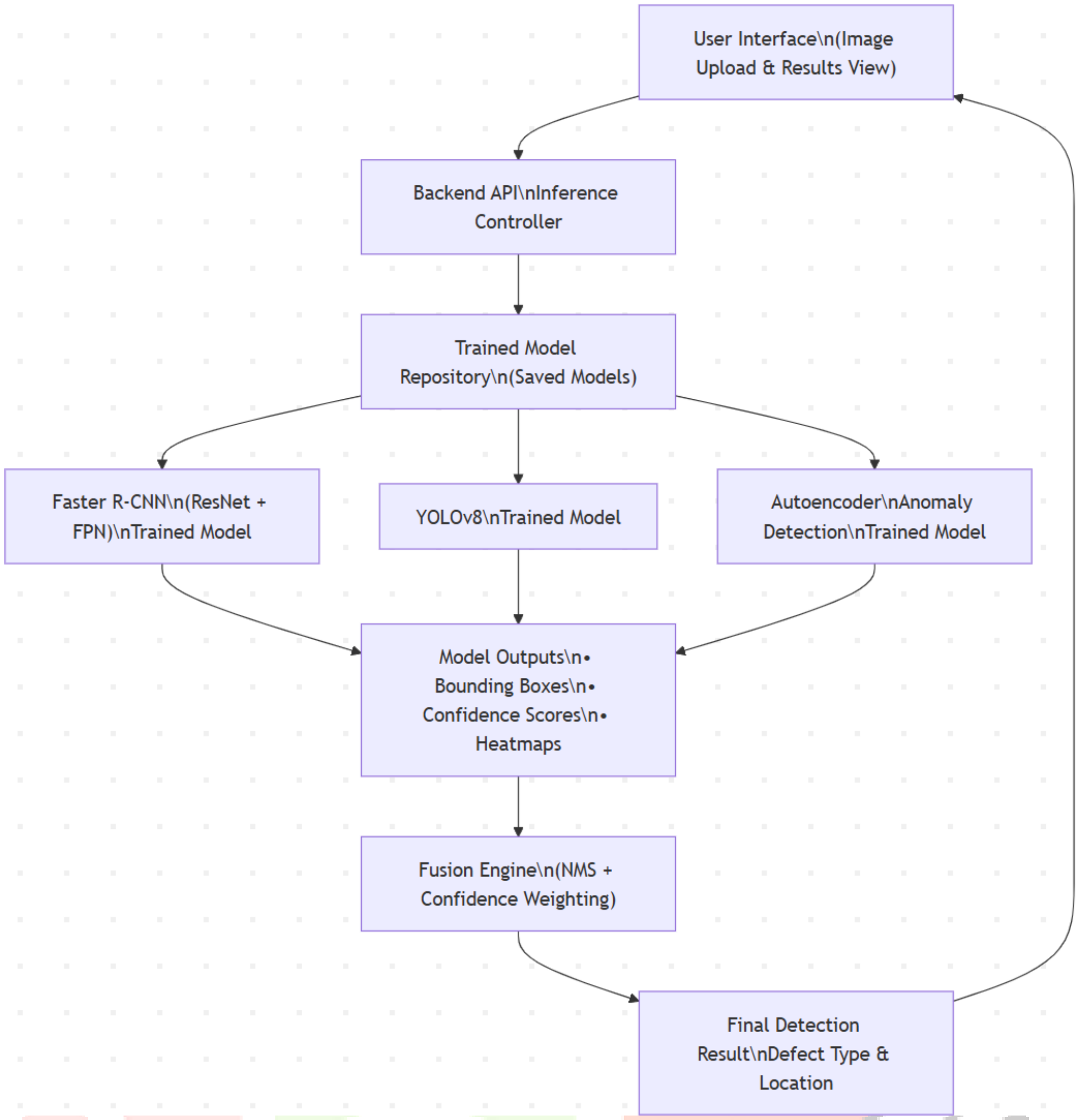

Fig. 2. Overall System Architecture of the Proposed Framework

E. Implementation Architecture

The implementation architecture of the proposed system illustrates the interaction between the user interface, backend inference controller, trained models, and fusion engine. The architecture emphasizes parallel model execution and intelli- gent fusion to enhance detection accuracy and reliability

F. Model Training

The model is trained using supervised learning with a multi-task loss function, which includes both classification and localization loss. Transfer learning with ResNet's pretrained weights is employed for faster convergence.

G. Testing and Inference

During inference, test images are passed through the trained network to obtain bounding boxes, class labels, and confidence scores.

H. Performance Evaluation

Precision, Recall, F1 score, and Mean Average Precision are used as performance evaluation metrics.

I. Implementation Details

The proposed system for image detection is developed using the PyTorch deep learning framework. A client-server architecture is used for efficient image detection. The user interface is responsible for image uploading and visualization.All input images are resized to a certain size and normalized. Data augmentation strategies such as random flip, scaling, and lighting are also applied to make it more robust in real-world industrial scenarios.

The backend inference controller dynamically loads pre- trained models from the model repository. The Faster R- CNN model utilizes a ResNet backbone integrated with a Feature Pyramid Network for multi-scale feature extraction and accurate localization. YOLOv8 is employed for real- time detection with low latency, while the Autoencoder-based model identifies anomalous regions by analyzing reconstruc- tion errors.

Inference is performed in parallel across all models to reduce processing time. Each model produces independent predictions, including bounding boxes, confidence scores, and heatmaps. These outputs are passed to a fusion engine, which applies Non-Maximum Suppression and confidence weighting to generate a single refined detection result. The final fused output, containing defect type and location, is sent back to the user interface for visual inspection. This implementation ensures efficient inference, improved detec- tion accuracy, and robust performance in complex industrial environments. .

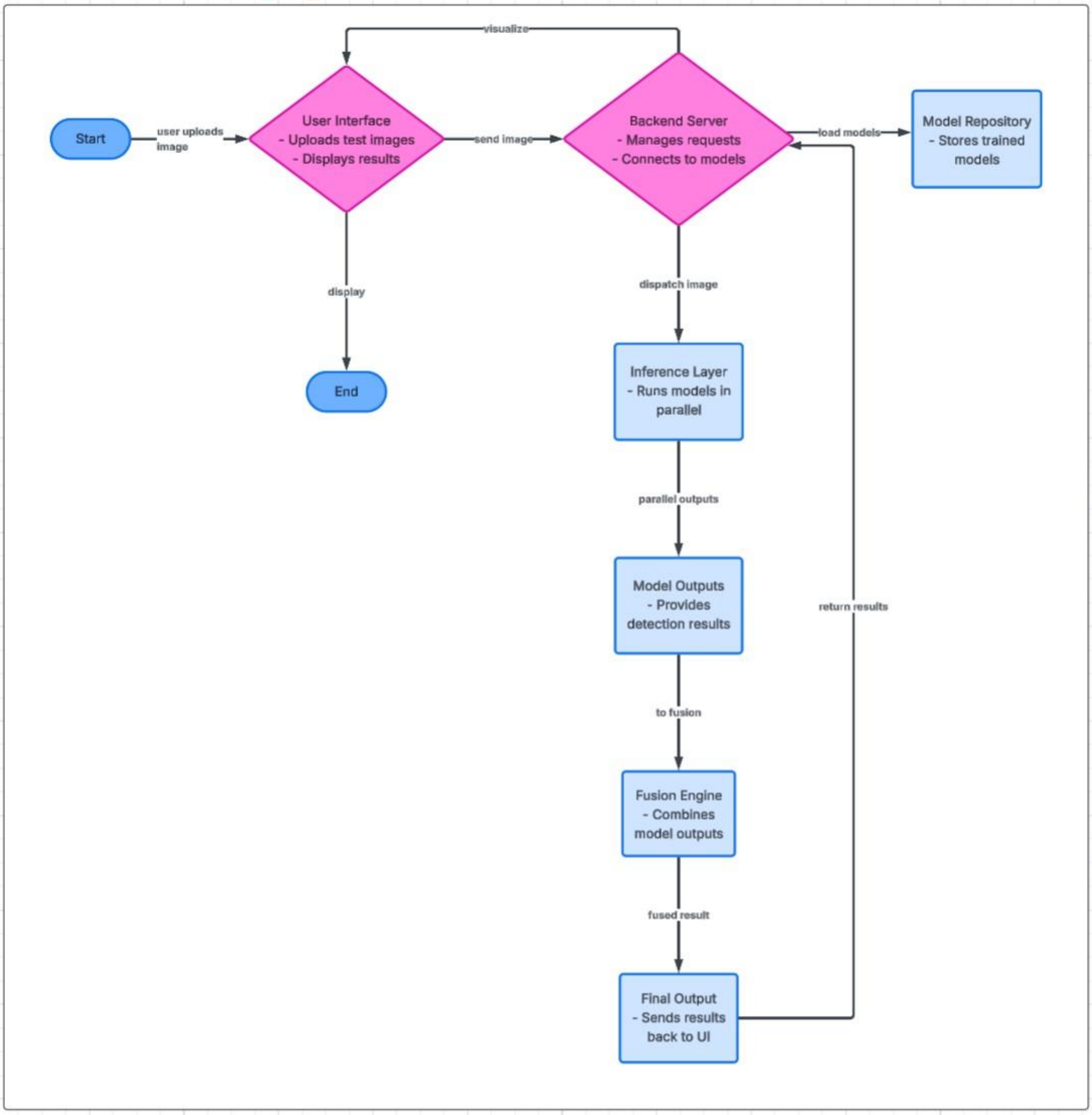

Fig. 3. Implementation Architecture of the Proposed Multi-Model Detection Framework

IV. RESULTS AND DISCUSSION

A. Training Convergence Analysis

The training convergence behavior of the proposed Faster R-CNN framework is illustrated in Fig. 4. A rapid decrease in loss is observed during the initial epochs, indicating effective learning of discriminative features. As training progresses, the loss reduction becomes gradual and stabilizes after later epochs, confirming stable convergence without oscillations or divergence. This behavior demonstrates that the adopted op- timization strategy successfully minimizes both classification and localization errors.

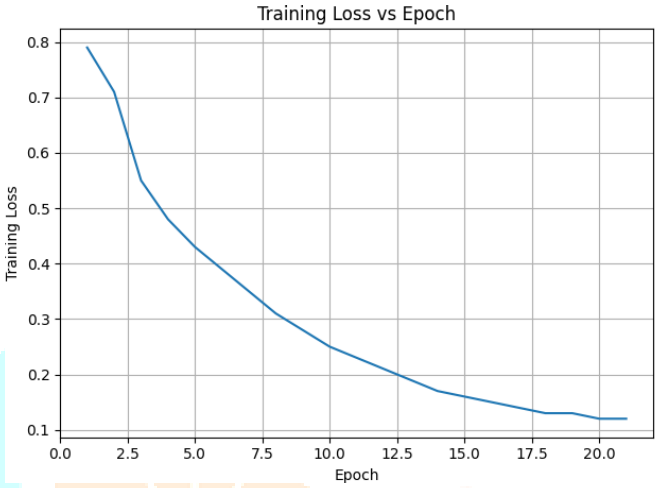

Fig. 4. Training loss variation across epochs

B. Validation Metric Analysis

Fig. 5 presents the progression of precision, recall, and F1-score across training epochs. All three metrics show a consistent upward trend, highlighting balanced learning be- havior. Precision improvement indicates a reduction in false positive detections, while recall enhancement demonstrates im- proved sensitivity towards actual foreign objects. The F1-score steadily increases and stabilizes in later epochs, confirming an effective trade-off between detection accuracy and robustness.

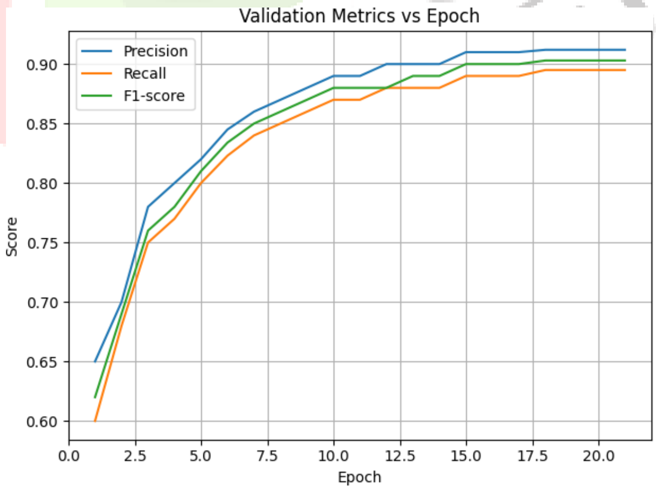

Fig. 5. Precision, recall, and F1-score progression during training

C. Confusion Matrix Analysis

The confusion matrix obtained on the validation dataset is shown in Fig. 6. Strong diagonal dominance is observed, indicating that most samples are correctly classified into their respective categories. Minimal off-diagonal values demon- strate low inter-class confusion, even among visually similar defect types. These results confirm the effectiveness of multi- scale feature extraction and region-based detection for accurate foreign object classification.

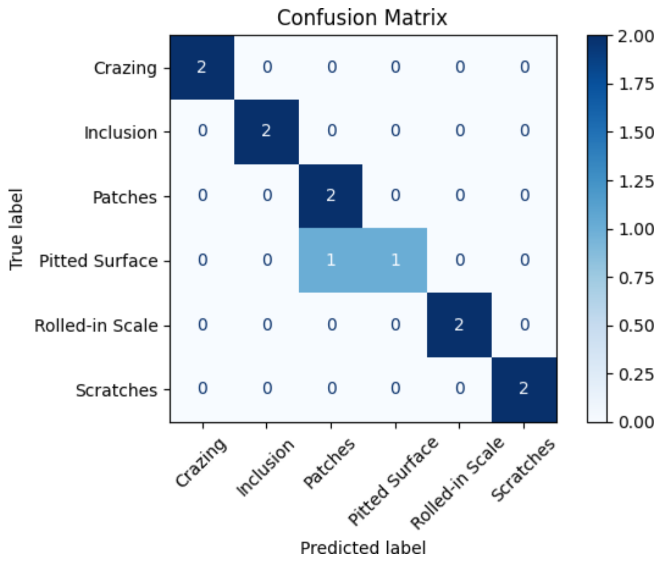

Fig. 6. Confusion matrix of the proposed detection framework

Overall, the experimental results confirm the robustness and reliability of the proposed approach for small foreign object detection in complex industrial environments.

V. CONCLUSION

The project is centered on enhancing the detection of small foreign objects in industrial images, a task that is often challenging due to variations in object size, lighting conditions, and background noise. Traditional image processing techniques and many existing deep learning models tend to struggle with maintaining high accuracy and robustness, particularly when dealing with very small or subtle defects. To address these limitations, the proposed system integrates a ResNet backbone with a Feature Pyramid Network (FPN), enabling effective multi-scale feature extraction and improving the model's ability to detect objects of varying sizes. Additionally, the use of Faster R-CNN ensures strong performance in both object detection and classification, contributing to more precise and reliable results. The primary goal of this approach is to achieve higher accuracy, minimize false negatives, and enhance overall efficiency. Ultimately, the system aims to support reliable, real-time quality control in industrial environments, helping to improve production standards and reduce the risk of undetected defects.

REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE* Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.

[2] R. Girshick, "Fast R-CNN," in Proc. IEEE Int. Conf. on Computer Vision (ICCV), pp. 1440–1448, 2015.

[3] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 580–587, 2014.

[4] K. He, G. Gkioxari, P. Dolla´r, and R. Girshick, "Mask R-CNN," in Proc. IEEE Int. Conf. on Computer Vision (ICCV), pp. 2961–2969, 2017.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 770–778, 2016.

[6] T.-Y. Lin, P. Dolla´r, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature Pyramid Networks for Object Detection," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2117–2125, 2017.

[7] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dolla´r, "Focal Loss for Dense Object Detection," in Proc. IEEE Int. Conf. on Computer Vision (ICCV), pp. 2980–2988, 2017.

[8] W. Liu, D. Anguelov, D. Erhan, et al., "SSD: Single Shot MultiBox Detector," in Proc. European Conf. on Computer Vision (ECCV), pp. 21–37, 2016.

[9] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-Scale Convolutional Neural Networks for Object Detection," in Proc. European Conf. on Computer Vision (ECCV), pp. 354–370, 2016.

[10] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable ConvNets v2: More Deformable, Better Results," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 9308–9316, 2019.

[11] Z. Li, C. Peng, G. Yu, X. Zhang, Y. Deng, and J. Sun, "Light-Head R-CNN: In Defense of Two-Stage Object Detector," arXiv:1711.07264, 2017.

[12] S. Zhang, L. Wen, X. Bian, Z. Lei, and S. Z. Li, "Single-Shot Refinement Neural Network for Object Detection," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 4203–4212, 2018.

[13] X. Xie, "A Review of Recent Advances in Surface Defect Detection using Computer Vision," IEEE Transactions on Instrumentation and Measurement, vol. 67, no. 10, pp. 2301–2318, 2018.

[14] D. Tsai and B. Hsiao, "Automatic Surface Inspection Using Wavelet Reconstruction," Pattern Recognition, vol. 34, no. 6, pp. 1285–1305, 2001.

[15] Y. Huang, C. Qiu, and Y. Guo, "Surface Defect Detection Using Deep Learning with Industrial Applications," IEEE Access, vol. 8, pp. 15384– 15397, 2020.

[16] S. Yang, L. Zhang, and J. Li, "Automatic Defect Detection on Steel Surfaces Using Deep Learning," Applied Sciences, vol. 9, no. 18, 2019.

[17] NEU Metal Surface Defect Dataset, "Northeastern University Surface Defect Database," Northeastern University, China.

[18] P. Bergmann, S. Lo¨we, M. Fauser, D. Sattlegger, and C. Steger, "Improving Unsupervised Defect Segmentation by Applying Structural Similarity," in Proc. British Machine Vision Conference (BMVC), 2018.

[19] T. Schlegl, P. Seebo¨ck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks," in Proc. Int. Conf. on Information Processing in Medical Imaging (IPMI), pp. 146–157, 2017.

[20] L. Ruff, R. Vandermeulen, N. Go¨rnitz, et al., "Deep One-Class Classi- fication," in Proc. Int. Conf. on Machine Learning (ICML), pp. 4393– 4402, 2018.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 779–788, 2016.

[22] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018.

[23] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv:2004.10934, 2020.