World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

Check for updates

A review of deep learning-based surface defect detection for castings

Xiaobin Liu * and Yantong Gong

School of Mechanical Engineering, Tianjin University of Technology and Education，Tianjin 300222.

World Journal of Advanced Research and Reviews, 2026, 30(03), 413-420

Publication history: Received on 27 April 2026; revised on 01 June 2026; accepted on 03 June 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.30.3.1608

Abstract

As a core technology in the high-end equipment manufacturing industry, casting requires crucial surface defect detection of products for quality control. Traditional nondestructive testing methods relying on manual experience present limitations including low efficiency, strong subjectivity, and poor adaptability, failing to meet modern industrial automation inspection requirements. Deep learning technology addresses the inefficiency of manually designed features in conventional approaches through convolutional neural networks and other models that automatically extract deep image features, enabling end-to-end defect detection. Current research primarily focuses on two-stage and single-stage object detection algorithms - the former emphasizes accuracy but suffers from computational complexity, while the latter prioritizes speed at the expense of small target detection capability. Domestic and international studies continue to optimize detection performance through network structure improvements, lightweight design implementation, and efficient loss function integration. Future research should enhance model generalization in complex scenarios, balance detection accuracy with operational speed, explore multi-modal data fusion and self-supervised learning technologies, thereby advancing casting defect detection systems toward intelligent and efficient development to provide technical support for industrial quality management.

Keywords: Deep Learning; Casting Surface Defect Detection; Object Detection Algorithms; Convolutional Neural Networks; Industrial Automated Inspection

1. Introduction

The high-end equipment manufacturing industry is a pillar of the national economy and a key driving force for industrial transformation and upgrading. It plays a vital role in promoting technological innovation and enhancing manufacturing competitiveness. As one of the fundamental and enabling technologies of modern equipment manufacturing, casting is widely employed in various sectors of the national economy, including aerospace, shipbuilding, mechanical and electronic engineering, and transportation. Consequently, the quality and reliability of castings have a direct impact on the performance, safety, and service life of industrial products.

According to their visual characteristics, casting defects can generally be classified into three categories: discrete point defects, linear defects, and area-based surface defects. Point defects may appear as isolated individual points, clustered distributions, or randomly scattered patterns. These defects include both regular and periodic features, such as roll marks, as well as irregular and non-directional defects, including accidental mechanical scars and localized corrosion. Linear defects are primarily manifested as elongated surface anomalies and can be further divided into continuous and discontinuous forms. Their orientations may extend longitudinally or transversely, among which casting cracks are the most representative type. Area-based defects refer to localized abnormal regions distributed over specific surface areas of a casting. Typical examples include oxidation marks, stain deposits, and other surface quality irregularities that adversely affect the appearance and structural integrity of cast products[1].

* Corresponding author: Yantong Gong

Among the various casting defects, porosity is one of the most significant and frequently encountered types. Porosity defects can be further categorized into several subtypes, including shrinkage porosity (micro-shrinkage, macro-shrinkage, and pipe shrinkage), blowhole porosity, and gas porosity. Such defects often lead to leakage problems, surface imperfections, and degradation of mechanical properties and fatigue performance [2]. The mechanical properties of final castings are primarily influenced by their chemical composition, melt quality, and secondary dendrite arm spacing (SDAS) [3]. In practical manufacturing processes, the presence of porosity defects significantly reduces the mechanical strength and fatigue resistance of materials. Furthermore, high-pressure die-cast products containing excessive porosity cannot be effectively subjected to heat treatment to enhance their mechanical properties. Therefore, accurately identifying and detecting defects in castings is of great importance for ensuring product quality and operational reliability [4–5].

Traditional non-destructive testing (NDT) methods for casting defect inspection mainly include radiographic testing (RT), ultrasonic testing (UT), magnetic particle testing (MPT), eddy current testing (ECT), and penetrant testing (PT) [6]. However, these conventional inspection techniques face several limitations in practical applications. For example, ultrasonic testing and magnetic particle testing require high-quality surface preparation and impose stringent requirements on the surface finish of the inspected components. Eddy current testing, on the other hand, suffers from equipment adaptability issues, as different coils must be frequently replaced to accommodate variations in the electromagnetic characteristics of different workpieces. More importantly, most existing inspection systems rely heavily on the experience and subjective judgment of operators. Consequently, the stability, consistency, and repeatability of inspection results are often affected by individual differences, resulting in limited objectivity during defect identification. Therefore, modern manufacturing industries require defect detection methods that can be seamlessly integrated into production processes while maintaining both high reliability and efficiency.

Machine vision-based inspection methods utilize industrial cameras and image-processing algorithms to detect casting defects. Although these methods have achieved considerable success, they still suffer from limited generalization capability and poor portability across different application scenarios [7]. The conventional machine vision-based defect inspection framework is illustrated in Fig. 1.

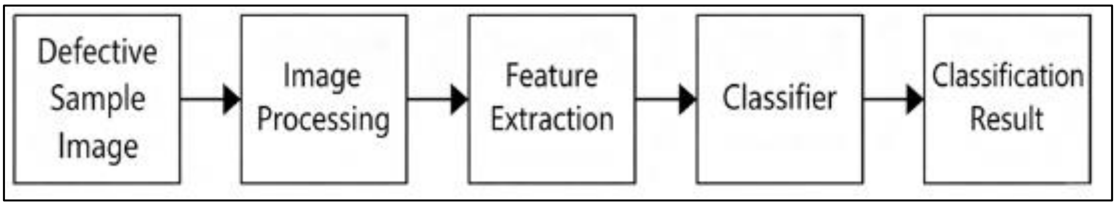

Figure 1 Conventional machine vision-based defect inspection process

With the rapid advancement of computer vision technology, an increasing number of researchers have applied deep learning techniques to surface defect detection. Compared with traditional machine vision approaches, deep learning methods are capable of extracting deeper and more abstract image features. Through operations such as convolution and pooling, deep learning models possess stronger feature representation capabilities and eliminate the need for manually designed feature extraction rules, thereby enabling end-to-end model development and automated defect recognition [8–9].

2. Research Progress in Casting Surface Defect Detection

2.1. Early Development of Machine Vision-Based Industrial Inspection

In the field of industrial automated inspection, developed Western countries started earlier than China and accumulated considerable technological advantages in both theoretical research and engineering applications. As early as the 1970s, with the further advancement of industrial automation, machine vision technology began to be introduced into industrial production in the United States. In the automotive manufacturing industry, for example, machine vision-based inspection systems effectively addressed several limitations of traditional manual quality inspection, such as low efficiency and poor consistency. Through automated image acquisition and intelligent image analysis, these systems significantly improved the efficiency, accuracy, and repeatability of component inspection, thereby providing an important technical foundation for the intelligent transformation of modern manufacturing systems [10].

With increasing attention paid to machine vision theory in developed countries, machine vision technology gradually gained widespread recognition and experienced rapid development. By the end of the twentieth century, it had been widely applied in various industrial fields. In the area of defect detection based on machine vision, several enterprises from developed countries had already developed mature solutions for practical industrial inspection tasks and designed corresponding machine vision-based defect detection systems. A representative example is Parsytec, a German company that developed a real-time inspection system capable of detecting various types of steel plate defects, including scratches, inclusions, and cracks, thereby enabling online automatic inspection in industrial production.

Studies related to the application of convolutional neural networks in defect detection are summarized in Table 1.

Table 1 Application of Convolutional neural networks (CNN) in materials science

Research Objective	Brief Conclusion	Reference
Mechanical property prediction	CNN outperforms standard shallow machine learning algorithms	[11]
Microstructure classification of steel	CNN outperforms standard shallow machine learning algorithms	[12]
Microstructure reconstruction and corresponding property prediction	CNN outperforms standard shallow machine learning algorithms	[15]
Microstructure segmentation of steel	Satisfactory results are obtained based on the CNN model	[14]
Defect detection of castings via X-ray inspection	The proposed defect detection system achieves state-of-the-art performance for casting defect detection	[15]
Investigation on microstructure and properties of photovoltaic materials	Satisfactory results are obtained based on the CNN model	[16]

2.2. Research Progress in China

From the perspective of manufacturing development, China has achieved continuous growth in industrial scale and has become one of the world's leading manufacturing economies. With the evolution of market demand and supply relationships, consumer expectations regarding product qualification rates have gradually shifted from basic functional verification to comprehensive quality control. This shift has directly promoted technological innovation in industrial inspection systems. By introducing machine vision-based intelligent recognition systems, traditional manual visual inspection can be transformed into automated, high-throughput, and highly accurate quality inspection, thereby establishing a new quality control paradigm for intelligent manufacturing [17–19].

With the development of these emerging technologies, machine vision theory has received increasing attention from both academia and industry in China. Reference [20] proposed an MFE-CasMVSNet network for steel plate surface defect detection. This method integrates point cloud processing techniques and improves three-dimensional reconstruction accuracy and information extraction capability through the PFEM and MFAFM modules. In addition, CSDBSCAN was introduced to achieve accurate defect identification and visualization of three-dimensional detection boxes. Experimental results on the NEU dataset showed that the recognition rate reached 95%, significantly improving detection and classification performance and providing an effective solution for quality control in steel production.

Reference [21] proposed YOLO-VDCW, a lightweight algorithm for strip steel surface defect detection. Based on YOLOv8, the algorithm incorporates VanillaNet, the C2fDSConv module, and a coordinate attention mechanism, while adopting the Wise-IoU loss function. Experimental results on the NEU-DET dataset demonstrated that YOLO-VDCW achieved a mean average precision of 79.8%, while reducing computational cost and parameter size by 34.1% and 36.8%, respectively. In addition, the detection speed was improved by 37.9%, indicating that the proposed method can effectively enhance both detection accuracy and computational efficiency and is suitable for practical industrial defect detection tasks.

In the field of image-based defect detection, deep learning algorithms constitute an important component of intelligent inspection systems, with hardware-based image processing modules serving as their implementation platform. Unlike mainstream non-destructive testing methods, deep learning-based approaches focus on deeply mining target features and emphasize the representational capability of neural networks. Therefore, their performance is less constrained by hardware conditions. Compared with traditional image processing methods, deep learning approaches substantially reduce the dependence on manually designed feature extractors. In conventional methods, feature extraction is usually designed manually, requiring data screening from scratch. Different feature extraction strategies and classifiers must be selected according to the specific characteristics of the data, and extensive trial-and-error experiments are often required to obtain methods that satisfy industrial requirements. This process is inefficient, costly, and highly application-specific. As a result, traditional image processing algorithms generally lack universality and transferability across different industrial scenarios.

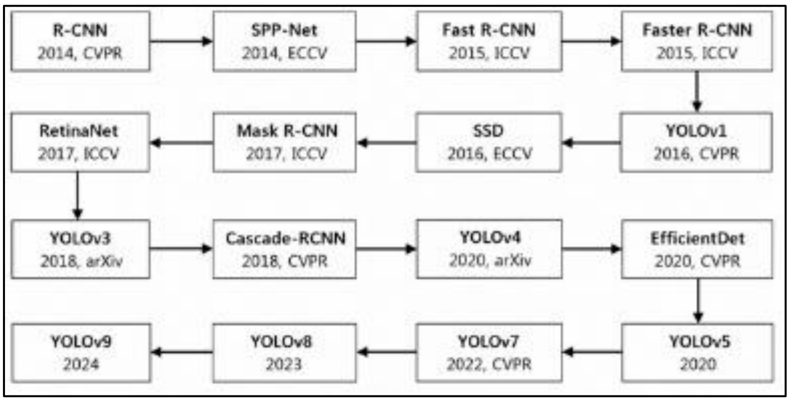

Figure 2 Development history of object detection algorithm

Fig. 2 illustrates the development history of object detection algorithms and clearly presents the evolutionary trajectory of related techniques. Deep learning-based object detection algorithms can generally be divided into two main categories. The first category consists of region-based two-stage object detection algorithms, such as R-CNN[22], Fast R-CNN[23], Faster R-CNN[24], and Mask R-CNN[25]. The second category includes regression-based one-stage object detection algorithms, such as YOLO[26] and SSD[27]. These two categories reflect continuous efforts by researchers to improve detection accuracy and detection speed, respectively [28–30]. The advantages and limitations of representative object detection network models are summarized in Table 2.

Table 2 The advantages and disadvantages of the target detection network model

Model	Category	Year	Network Architecture	Advantages	Limitations
RCNN	Two-stage	2014	AlexNet	Introduces region proposals and CNNs, establishing a new two-stage object detection framework.	High model complexity, slow training speed, and low computational efficiency.
SPP-Net	Two-stage	2014	ZFNet	Enables multi-scale input images.	The convolutional layer parameters cannot be updated through backpropagation.
Fast RCNN	Two-stage	2015	VGG	Introduces the ROI Pooling layer.	Cannot meet the requirements of real-time applications.
Faster RCNN	Two-stage	2015	VGG	Improves detection accuracy and speed, and performs well in small-object detection.	Cannot meet the requirements of real-time applications.
Yolov1	One-stage	2016	24-layer 卷积层	Proposes a new one-stage object detection framework.	Shows relatively low detection performance for small or irregularly shaped defects.

SSD	Two-stage	2016	VGG	Achieves relatively high accuracy and fast detection speed.	Has difficulty detecting small defect targets.
Mask R-CNN	Two-stage	2017	ResNet	Enables high-precision instance segmentation.	The model is complex and requires substantial computational resources.
RetinaNet	One-stage	2017	FPN	Improves the accuracy of one-stage object detectors.	Addresses the imbalance between positive and negative samples.
Yolov2	One-stage	2017	DarkNet-19	Uses prior boxes and feature fusion strategies to improve detection efficiency.	Computationally complex and difficult to apply to small-defect detection.
Yolov3	One-stage	2018	DarkNet-53	Supports multi-label prediction through independent logistic regression.	Has relatively low recall and limited detection performance under occlusion and crowded conditions.
Cascade R-CNN	Two-stage	2018	RPN	Achieves high accuracy through multi-stage optimization.	Computationally complex and difficult to apply to small-defect detection.
Yolov4	One-stage	2020	CSPDarkNet	Integrates multiple methods and models, resulting in relatively high detection efficiency	Has a high false detection rate for predicted bounding boxes.
EfficientDet	One-stage	2020	BiFPN	Efficient and highly accurate.	Suitable for resource-constrained environments.
Yolov5	One-stage	2020	CSPDarkNet	Increases the depth and width of the network.	Has difficulty detecting small defect targets.
Yolov7	One-stage	2022	CBS LAN MP-1	Enables real-time object detection.	The network architecture is relatively complex and has difficulty detecting small defect targets.
Yolov8	One-stage	2023	Darknet53	Provides more accurate and flexible detection results.	detection results. The model is relatively complex and requires substantial computational resources and training time
Yolov9	One-stage	2024	GELAN	Its innovative architecture enables more efficient utilization of model parameters.	Further optimization is still required due to structural complexity, limited scalability, and high training cost.

3. Conclusion

Surface defect detection for castings, which are core basic components in high-end equipment manufacturing, plays an essential role in ensuring the quality and reliability of industrial products. Traditional non-destructive testing methods have long been limited by their dependence on manual experience, low inspection efficiency, and insufficient adaptability to complex industrial environments. These limitations make conventional inspection paradigms increasingly difficult to meet the quality control requirements of intelligent manufacturing. The rapid development of deep learning has provided new opportunities for casting surface defect detection. By constructing intelligent models such as convolutional neural networks, deep learning-based methods can autonomously extract deep features from casting surface images and perform end-to-end defect detection. This fundamentally alleviates the subjectivity and inefficiency associated with manual feature extraction and has significantly improved the performance of defect detection systems.

From the perspective of technological evolution, current research mainly focuses on the coordinated development of two-stage and one-stage object detection algorithms. Two-stage algorithms represented by Faster R-CNN rely on region proposal networks (RPNs) and refined feature processing mechanisms, maintaining strong advantages in terms of detection accuracy. However, their high computational complexity and limited real-time performance restrict their application in industrial scenarios. In contrast, one-stage algorithms, particularly the YOLO series, have achieved substantial improvements in detection speed through innovative regression-based mechanisms. Nevertheless, these methods still face challenges in extracting discriminative features from small-scale defects. To further improve detection performance, researchers have explored multi-dimensional technical integration strategies, including feature pyramid fusion modules and lightweight network architectures at the model level, improved loss functions such as Wise-IoU at the algorithmic level, and multimodal information fusion strategies that integrate three-dimensional point clouds with two-dimensional images at the data level. Representative methods such as YOLO-VDCW demonstrate that architectural optimization can effectively reduce computational cost while improving detection speed, thereby providing feasible technical support for real-time industrial inspection systems.

Future research on casting defect detection should focus on breakthroughs in three key directions. First, robust model architectures with strong generalization capability should be developed to address complex industrial conditions, such as illumination variations, background interference, and heterogeneous defect scale distributions. Second, lightweight neural network design should be integrated with dynamic resource scheduling to establish a joint optimization mechanism that balances detection accuracy and inference speed, thereby satisfying the diverse requirements of discrete manufacturing scenarios. Third, further efforts are needed to overcome the bottlenecks of multi-source heterogeneous data fusion by integrating visible-light images, ultrasonic testing data, X-ray images, and other multimodal information. In addition, self-supervised learning frameworks can be introduced to alleviate the scarcity of annotated defect data and support the construction of intelligent inspection systems with autonomous learning and adaptive optimization capabilities.

In the broader context of the integration of Industry 4.0 and artificial intelligence, the continuous advancement of casting defect detection technologies will provide strong support for upgrading quality control systems in high-end equipment manufacturing. It will also contribute to the development of more reliable, efficient, and intelligent industrial inspection systems.

Compliance with ethical standards

Acknowledgments

This work was supported by the School of Mechanical Engineering, Tianjin University of Technology and Education. The authors thank the institution for the academic and resource support for completing this review work.

Disclosure of conflict of interest

The authors declare that they have no conflict of interest for this published manuscript.

No conflict of interest to be disclosed.

Statement of ethical approval

This article is a review paper based on existing published literatures, and no human or animal experimental research is carried out, thus ethical approval is not needed.

Statement of informed consent

No human subjects are enrolled in this research, so informed consent is not applicable to this manuscript.

References

[1] Lin J,Yao Y, Ma L,et al.Detection of a casting defect tracked by deep convolution neural network[J].International Journal of Advanced Manufacturing Technology,2018, 97(4):1-9.

[2] C. Pieringer and D. Mery.,“Flaw detection in aluminium die castings using simultaneous combination of multiple views.,”Insight-Non-Destructive Test. Cond. Monit., vol. 52, no. 10, pp. 548–552, 2010.

[3] D. Banerjee, V. Kukreja, S. Rana, M. Manwal, and S. Hariharan., “Enhancing Snake Plant Disease Classification through CNN-Random Forest Integration.,” in In 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), 2023, pp. 1–6.

[4] D. Banerjee, V. Kukreja, A. Verma, N. Garg, and S. Hariharan., “Hybrid CNN & Random Forest Model for Effective Guava Leaf Disease Diagnosis.,” in In 2023 International Conference on Research Methodologies in Knowledge Management, Artificial Intelligence and Telecommunication Engineering (RMKMATE), 2023, pp. 1–6.

[5] A. Kaur, V. Kukreja, D. Banerjee, and K. Madan., “MintLeafNet: ACNN Model for Accurate and Efficient Detection of Multiple Mint Leaf Diseases.,” in In 2023 4th IEEE Global Conference for Advancement in Technology (GCAT), 2023, pp. 1–5.

[6] D. Banerjee, V. Kukreja, A. Gupta, V. Singh, and T. P. S. Brar., “Combining CNN and SVM for Accurate Identification of Ridge Gourd Leaf Diseases.,” in In 2023 3rd Asian Conference on Innovation in Technology (ASIANCON), 2023, pp. 1–5.

[7] D. Banerjee, V. Kukreja, S. Hariharan, V. Jain, and V. Jindal., “Predicting Tulip Leaf Diseases: A Integrated CNN and Random Forest Approach.,” in In 2023 World Conference on Communication & Computing (WCONF), 2023, pp. 1–6.

[8] R. Gupta, V. Anand, S. Gupta, and D. Koundal., “Deep learning model for defect analysis in industry using casting images.,” Expert Syst. Appl., vol. 232, p. 120758, 2023.

[9] M. Habibpour et al., “An uncertainty-aware deep learning framework for defect detection in casting products.,” arXiv Prepr. arXiv2107, p. 11643, 2021.

[10] L. Duan, K. Yang, and L. Ruan., “Research on automatic recognition of casting defects based on deep learning.,” IEEE Access, vol. 9, pp. 12209–12216, 2020.

[11] C. Herriott, A.D. Spear, Predicting microstructuredependent mechanical properties in additively manufactured metals with machine-and deep-learning methods. Comput. Mater. Sci. 175, 109599 (2020).

[12] S.M. Azimi, D. Britz, M. Engstler, M. Fritz, F. Mu¨cklich, Advanced steel microstructural classification by deep learning methods. Sci. Rep. 8(1), 1–14 (2018).

[13] X. Li, Y. Zhang, H. Zhao, C. Burkhart, L.C. Brinson, W. Chen, A transfer learning approach for microstructure reconstruction and structure-property predictions. Sci. Rep. 8(1), 1–13 (2018).

[14] B.L. DeCost, B. Lei, T. Francis, E.A. Holm, High throughput quantitative metallography for complex microstructures using deep learning: a case study in ultrahigh carbon steel. arXiv preprint arXiv:1805. 08693 (2018).

[15] M.K. Ferguson, A. Ronay, Y.-T.T. Lee, K.H. Law, Detection and segmentation of manufacturing defects with convolutional neural networks and transfer learning. Smart Sustain. Manuf. Syst. 2, 137–164 (2018).

[16] B.S.S. Pokuri, S. Ghosal, A. Kokate, S. Sarkar, B. Ganapathysubramanian, Interpretable deep learning for guided microstructure property explorations in photovoltaics. Comput. Mater. 5(1), 1–11 (2019).

[17] A. Agrawal, P.D. Deshpande, A. Cecen, G.P. Basavarsu, A.N. Choudhary, S.R. Kalidindi, Exploration of data science techniques to predict fatigue strength of steel from composition and processing parameters. Integr. Mater. Manuf. Innov. 3(1), 8 (2014).

[18] D. Cica, D. Kramar, Intelligent process modeling and optimization of porosity formation in high-pressure die casting. Inter. Metalcast. 12(4), 814–824 (2018).

[19] F. Nikolic, I. Stajduhar, M. Canadija, Aluminum Microstructure Inspection Using Deep Learning: a Convolutional Neural Network Approach Toward Secondary Dendrite Arm Spacing Determination. in 4th Edition of My First Conference, University of Rijeka, Faculty of Engineering, Rijeka, Croatia (2020).

[20] Lan H., Yu J. B. Steel plate surface defect detection based on deep learning and three-dimensional reconstruction[J]. Journal of Zhejiang University (Engineering Science), 2023, 57(3): 466–476, 561.

[21] Liu F. C., Zhang J., Xue T., et al. YOLO-VDCW: A novel lightweight algorithm for strip steel surface defect detection[J/OL]. China Metallurgy, 1–13 [2024-06-20].

[22] 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.

[23] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[24] 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, 28-36.

[25] He K, Gkioxari G, Dollár P, et al. Mask rcnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.

[26] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[27] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer VisionECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[28] B. Li, W. Jiang, J. Gu, et al., “A review of object detection algorithms based on convolutional neural networks,” Computer & Digital Engineering, vol. 50, no. 5, pp. 1010–1017, 2022.

[29] G. Sharma, V. Anand, R. Chauhan, H. S. Pokhariya, S. Gupta, and G. Sunil., “Advanced Deep Learning Methods for Comprehensive Quality Inspection and Fault Detection in Casting Processes.,” in In 2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT), 2024, pp. 602–606.

[30] D. Mery, “Aluminum casting inspection using deep learning: a method based on convolutional neural networks.,” J. Nondestruct. Eval., vol. 39, no. 1, p. 12, 2020.