A Study of Computational Vision Methods for Corrosion Detection in Industrial Assets

Anderson Gonçalves Marco*, Marcos de Souza Oliveira*
Rodrigo Leite Prates*
Murilo Cruz Lopes*
Rogério Guedes Casal*
Cristiani Vilela Ribeiro Guimarães*
*Inatel Competence Center, MG, Brazil

***Abstract*—Corrosion has a significant impact on the global economy, leading to high costs for preventing damage, repairing deterioration, and replacing defective components. Industry 4.0 brings computational vision as a key role for visual inspection of assets ensuring quality and reducing the downtime. This study conducts a comparative analysis of computer vision techniques for detecting corrosion, with an emphasis on segmentation neural networks built upon U-Net architectures, focusing on the trade-off between segmentation accuracy and computational efficiency. We benchmark four U-Net-based segmentation models using pre-trained backbones: MobileNetV2, ResNet152, InceptionV4 and VGG16 on a publicly available data set of corroded metallic surfaces. All models are trained using the full dataset and evaluated based on per-image F1-score and training time. Our results show that InceptionV4 achieves the highest F1-score (0.895), while MobileNetV2 offers a comparable score (0.892) with nearly half the training time. These findings suggest that lightweight backbones can deliver accurate segmentation with significantly reduced computational cost, making them promising candidates for near real-time industrial inspection systems under resource constraints.**

I. Introduction

Corrosion represents a significant economic burden globally. In the United States, a landmark report titled *"Corrosion Cost and Preventive Strategies in the United States"* [1] estimated that corrosion-related expenses accounted for 4.2% of the gross national product (GNP) in 1975 and 3.1% in 1998—figures that, for comparison, approached the country's public spending on education in the same year (4.8% of GNP)[1]. Considering more recent studies, it's important to highlight that the problem remains critical on a global scale. According to the Association for Materials Protection and Performance (AMPP), corrosion-related costs now exceed US$2.5 trillion annually, equivalent to approximately 3.4% of global GDP [2], with U.S. losses alone surpassing US$450 billion per year. Iannuzzi and Frankel [3] similarly estimate that corrosion impacts account for 3% to 4% of global GDP, reinforcing the need for effective mitigation strategies.

In response to this persistent challenge, international organizations have intensified awareness and advocacy efforts. One example is the World Corrosion Awareness Day 2025, promoted by AMPP in collaboration with the World Corrosion Organization (WCO) and the European Federation of Corrosion (EFC). These initiatives emphasize corrosion as a widespread and costly threat to infrastructure, safety, and the global economy[2]. Such urgency calls for innovative computational approaches capable of early and accurate detection of corrosion-related defects.

Visual defect detection plays a pivotal role in industrial quality inspection, including for the detection of corrosion ([4] and [5]), ensuring product reliability and reducing manufacturing costs [6]. With the advancement of Industry 4.0, deep learning-based visual inspection systems have increasingly supplanted traditional rule-based approaches due to their robustness and adaptability [7]. Among these, convolutional neural networks (CNNs), particularly U-Net and its variants, have shown strong performance in the semantic segmentation of surface defects in various manufacturing domains, including metal processing [8], PCB soldering [9], steel plate inspection [10], and Corrosion defects [11], that present unique challenges due to their high visual variability and the need for precise localization to inform effective maintenance strategies.

To enhance segmentation performance, recent works have proposed architectural refinements to the U-Net framework, including the integration of advanced backbones such as ResNet and EfficientNet [12]. These adaptations have led to measurable gains; for example, EffUNet-B7 achieved high performance in the detection of metal surface defects [13]. Currently, hybrid models that combine segmentation with decision-making modules have been explored to improve performance in annotation-limited or few-shot scenarios [14].

Despite these advances, important questions remain. The impact of backbone complexity, annotation granularity, and dataset size is still underexplored in constrained industrial environments, where labeled data are scarce and inference must occur in near real time [15]. Moreover, many evaluations are based on large-scale or manually curated datasets that do not reflect the limitations of small-batch defect inspection settings. As an alternative to address the adaptability challenges

[1]https://tradingeconomics.com/united-states/public-spending-on-education-total-percent-of-gdp-wb-data.html

[2]https://www.ampp.org/blogs/webmasternaceorg/2025/04/22/global-campaign-urges-action-on-corrosion-crisis

of complex industrial scenarios and the lack of high-quality data, the adoption of simulators and digital twins has been increasingly considered [16]. The core idea is to create a digital representation, often using 3D modeling, of the physical world in order to anticipate and mitigate natural variations and implementation issues, enabling a more robust design phase before deployment.

Given the recent availability of publicly accessible image datasets of corrosion-related defects, this study instead focuses on a comparative analysis of U-Net variants employing different backbone architectures under data-limited conditions. Our focus is on segmenting corrosion-related surface defects in industrial components, a task of practical importance in sectors such as automotive, infrastructure, and manufacturing. In doing so, we analyze the trade-offs between segmentation accuracy and computational cost, as well as the minimal annotation requirements for effective training.

While U-Net has been predominantly employed in the medical imaging domain, our work extends its application to the industrial inspection of corrosion-related defects, a domain with substantially different visual characteristics, defect morphologies, and data availability constraints. This cross-domain adaptation represents a novel contribution, as it evaluates the robustness and generalization capacity of U-Net-based architectures under challenging, non-medical imaging conditions. By systematically exploring multiple backbone configurations, we provide insights that bridge the gap between established segmentation methodologies and emerging needs in industrial quality assurance.

Although the use of digital twins is not directly explored in this work, the findings derived from the analysis of U-Net backbones are intended as a foundational step toward future applications within simulated environments and digital twin frameworks.

The remainder of this paper is structured as follows: Section II reviews related work; Section III details the experimental setup and presents results; and Section IV concludes the study and outlines future research directions.

II. Related Works

The application of computer vision techniques for defect detection in industrial settings has been extensively explored [17], [18]. However, an increasingly prominent approach involves the use of image segmentation methods. In addition to simply classifying whether a piece or product is defective or not, segmentation methods provide crucial information by identifying the exact location or region of the defect (region of interest). In this context, the U-Net model has made a considerable contribution to image segmentation in recent years [19].

U-Net is a convolutional neural network architecture that comprises a contracting path (encoder), responsible for extracting representative features from the input image, and an expansive path (decoder), which enables precise segmentation by reconstructing spatial information. The selection and modification of the U-Net backbone, i.e. the feature extraction component, has been the focus of numerous studies. For example, [20] proposes a modified U-Net backbone to address change detection in remote sensing images. Similarly, [10] explores different U-Net-based architectures for defect segmentation on metallic surfaces.

Several works have investigated improvements to the original U-Net architecture to address domain-specific challenges in industrial inspection. In [13], the authors apply U-Net to detect scratches and surface anomalies in metal sheets, using a real-world production dataset. In a similar context, Leontaris et al. [8] investigate multiple U-Net variants trained on controlled datasets for metallic surface inspection. Li and Liu [9] propose architectural modifications, such as the use of a dynamic hybrid attention module and separable convolutions, for precise segmentation of soldering defects in PCBs. Ahmed et al. [14] present a segmentation-based deep learning pipeline that also focuses on robustness across different defect types.

Moreover, hybrid approaches are frequently adopted in the state-of-the-art for industrial defect detection. These typically involve two stages: first, feature extraction is performed using well-established deep learning architectures such as convolutional networks; then, classical classification algorithms are applied to the extracted features. Studies such as [21] and [22] exemplify the effectiveness of this hybrid strategy in various industrial defect detection scenarios.

Overall, the growing body of literature emphasizes not only the effectiveness of U-Net and its variants in industrial inspection but also the potential for optimization through backbone selection, lightweight modules, and training under real-world constraints such as limited data or small defect areas.

III. Experimental Results

This section presents the experimental design adopted to evaluate the performance of U-Net-based architectures for surface defect segmentation in industrial components. Our goal is to analyze how different backbone complexities influence the trade-off between accuracy and computational cost. To ensure a systematic and reproducible evaluation, we organize this section into four parts.

Subsection III-A describes the dataset used in our experiments, including its characteristics, preprocessing steps, and the nature of the annotated defects. Subsection III-B details the U-Net variants considered, highlighting the architectural differences between the backbone networks. Subsection III-C outlines the experimental protocol, including data splits, training settings, evaluation metrics, and hardware specifications. Finally, Subsection III-D presents and discusses the results obtained, focusing on segmentation performance, model efficiency, and the implications for near real-time defect inspection in constrained industrial scenarios.

A. Dataset

To evaluate the proposed segmentation models, we selected a publicly available dataset from the Roboflow platform entitled “cor Dataset” [23]. This dataset was chosen due to its strong alignment with the problem of detecting corrosion in industrial components, offering a large volume of annotated data and diverse visual conditions. It comprises 4,949 RGB images with a resolution of 640×640 pixels, each annotated with pixel-level segmentation masks. The images depict various types and severities of corrosion across a broad range of metallic surfaces, including car parts, screws, chains, plates, pipes, and bars (see Figure 1). This diversity enhances the generalization of the models trained on the dataset and better reflects real-world inspection scenarios. Annotations are provided in the COCO segmentation format and demonstrate a high level of consistency and detail. For model training and evaluation, the dataset was randomly divided into three subsets: 70% for training, 15% for validation, and 15% for testing. This split allows for a balanced assessment of model performance and generalization. The combination of scale, variety, and annotation quality makes this dataset particularly well suited for benchmarking segmentation models in industrial corrosion detection contexts.

(a) Car (b) Metal Plate (c) Metal Bolt

Fig. 1: Examples of annotated corrosion images from the dataset, showing different surface types and severity levels.

B. Unet Architectures

U-Net is a convolutional neural network architecture originally proposed for biomedical image segmentation. Its structure is composed of a contracting path (encoder), which captures hierarchical features, and an expansive path (decoder), which reconstructs spatial resolution to generate a segmentation map. One of U-Net's main advantages is its ability to perform precise localization using relatively few training images, making it suitable for industrial inspection tasks under limited data availability.

In this study, we evaluate four U-Net variants based on different encoder backbones: ResNet152 [24], VGG16 [25], InceptionV4 [26], and MobileNetV2 [27]. These backbones were selected based on findings reported in the literature and to ensure variation in terms of network depth and number of parameters, allowing for a comparative analysis across architectures with different levels of complexity and computational demand.

- ResNet152-UNet introduces deep residual connections that facilitate the training of very deep networks and capture rich hierarchical features, albeit with higher computational cost.
- VGG16-UNet uses a simpler and widely adopted convolutional backbone with uniform layer configurations, serving as a strong baseline in many segmentation benchmarks.
- InceptionV4-UNet leverages multi-scale feature extraction within inception modules, which is particularly useful for capturing patterns of varying size and texture, such as corrosion.
- MobileNetV2-UNet provides a lightweight alternative with significantly fewer parameters, relying on depthwise separable convolutions, making it more suitable for near real-time or resource-constrained environments.

C. Methodology

All images used in the experiments were resized to a resolution of 224×224 pixels. This choice was based on a convention widely adopted in the literature for image classification and segmentation tasks [25], particularly in works that employ transfer learning [28]. Most of the pretrained encoder backbones evaluated in this study, such as ResNet152, VGG16, InceptionV4, and MobileNetV2 were originally trained on the ImageNet dataset [29], which uses 224×224 as the standard input size. Maintaining this resolution ensures compatibility with pretrained weights. Additionally, several recent works have adopted this same input size in segmentation pipelines for industrial applications, including [8], [10], [14], [30].

The training procedure begins by partitioning the dataset into three subsets: 70% for training, 15% for validation, and 15% for testing. The training and validation sets are employed during model optimization to update network weights and monitor generalization performance. All models are trained for 150 epochs using identical data splits and hyperparameters to ensure fair comparison across backbone architectures. The test set is held out for final evaluation, allowing the computation of performance metrics on previously unseen data.

To guide the optimization process, the Dice loss [31] is adopted as the objective function. Dice loss is particularly suitable for segmentation tasks involving imbalanced classes, as it emphasizes spatial overlap between predicted and ground-truth masks rather than individual pixel-wise accuracy. This property is essential in surface defect segmentation, where defective regions often occupy a small fraction of the image.

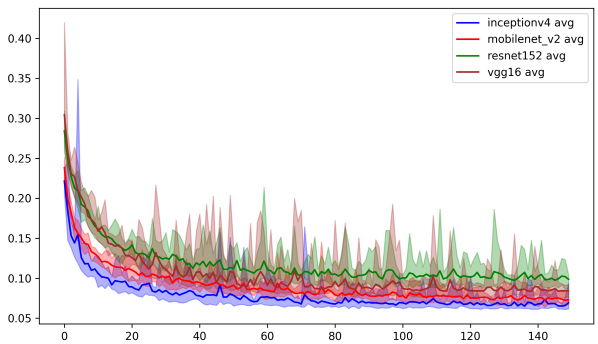
Model performance is assessed using the F1-score [32] and the Intersection over Union (IoU) [33], both widely adopted in semantic segmentation benchmarks. The F1-score provides a harmonic mean between precision and recall, offering a balanced view of classification performance in pixel-level predictions. Meanwhile, the IoU metric quantifies the spatial agreement between predicted and true masks by computing the ratio of their intersection over union. Higher IoU values reflect more accurate localization and shape estimation of defects.

D. Results

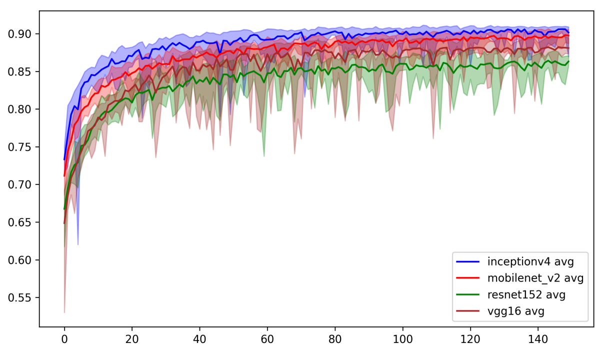
Our initial analysis is based on the results obtained during the training phase of the models. At this stage, we first examined the loss values in the training epochs, from which we observed that the MobileNetV2 and ResNet152 backbones exhibit a certain degree of instability in the values of the

TABLE I: Comparison of U-Net encoder backbones used in this study. FLOPs estimated for 224×224×3 input. "Params" = parameters. InceptionV4 depth is approximate due to its modular structure, where each block contains multiple layers.

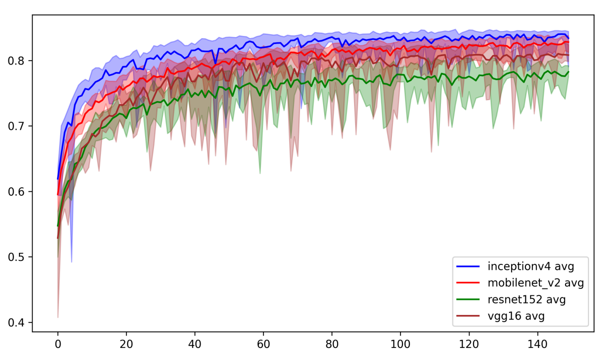
Backbone	**Params (M)**	**Depth**	**FLOPs (G)**	**Train-Time(Min)**	**Notes**
ResNet152	$\sim$60.2	152	$\sim$11.5	67.747 ± 4.898	Deep, robust features
VGG16	$\sim$138.4	16	$\sim$15.3	62.820 ± 0.292	Classic baseline
InceptionV4	$\sim$42.6	$\sim$130	$\sim$12.3	45.419 ± 0.048	Multi-scale filters
MobileNetV2	$\sim$3.5	53	$\sim$0.3	32.153 ± 0.120	Lightweight, fast



(a) Loss (dice)



(b) F1 score



(c) IoU

Fig. 2: Results of the benchmark training

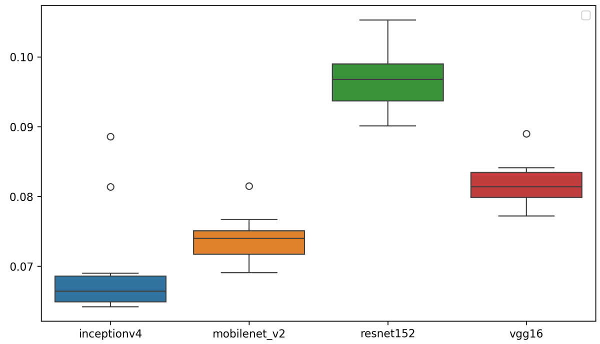
Dice coefficient (see Figures 2a and 3a). Another important observation is that, in absolute terms, the InceptionV4 backbone stands out as the best-performing option among the compared models, particularly when considering the F1-score (see Figures 2b and 3b) and IoU metrics (see Figures 2c and 3c). Table I and Table II presents the training time and mean with standard deviation of the F1 score per image in ten runs for each variant U-Net and Figure 3 the statistical results of the metrics. Among the evaluated models, InceptionV4 achieved the highest mean F1-score (0.895 ± 0.011), followed closely by MobileNetV2 (0.892 ± 0.004). Despite their similar segmentation performance, MobileNetV2 exhibited a lower average training time (32.15 ± 0.12 minutes) compared to InceptionV4 (62.82 ± 0.29 minutes), offering a favorable trade-off between accuracy and computational efficiency.

TABLE II: Mean and standard deviation of per-image F1-score and Iou over 10 runs.

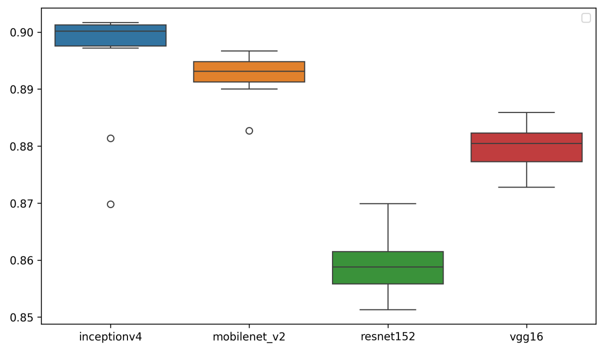
Backbone	**F1-Score**	**IoU**
ResNet152	0.859 ± 0.005	0.779 ± 0.006
InceptionV4	0.895 ± 0.010	0.828 ± 0.014
VGG16	0.880 ± 0.004	0.807 ± 0.005
MobileNetV2	0.892 ± 0.004	0.822 ± 0.005

In addition to the analysis of absolute average values, we applied the Tukey HSD test [34] to statistically compare the segmentation performance across different U-Net backbones. Table IV presents the results of this analysis based on the F1-scores obtained from ten independent simulations. The test confirms that the performance differences between InceptionV4 and ResNet152 and VGG16 are statistically significant ($p < 0.001$), as are the differences between MobileNetV2 and the same two models. These results suggest that lighter or more recent architectures, such as InceptionV4 and MobileNetV2, offer advantages in data-limited segmentation scenarios, likely due to a more balanced trade-off between depth, parameter count, and representational efficiency.

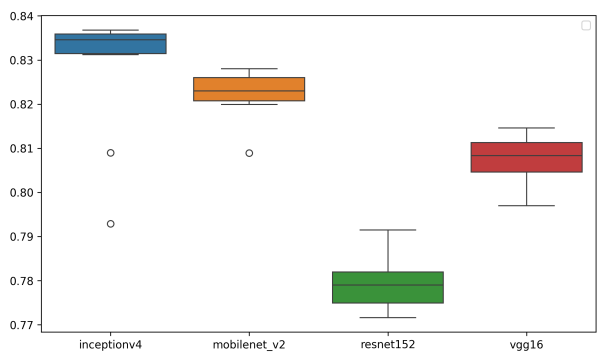
Furthermore, no statistically significant differences were observed between InceptionV4 and MobileNetV2 ($p = 0.7687$), indicating that both models show comparable performance in terms of segmentation precision. This similarity may arise from their architectural design choices, such as the use of depth-wise separable convolutions in MobileNetV2 and factorized convolutions in InceptionV4, which promote efficient feature extraction while maintaining high accuracy. These findings reinforce the importance of architectural efficiency in constrained environments, where overparameterized models



(a) Loss (dice)



(b) F1 score



(c) IoU

Fig. 3: Results of the benchmark training

like ResNet152 and VGG16 may suffer from overfitting or limited generalization due to the scarcity of labeled training data.

IV. Conclusion

This study presented a comparative analysis of U-Net architectures employing different backbone networks for the segmentation of corrosion-related surface defects under data-limited industrial conditions. Using a publicly available dataset, we evaluated the trade-offs between segmentation accuracy, computational complexity, and annotation efficiency. Factors that are particularly relevant in real-world inspection scenarios where labeled data are scarce and near real-time inference is required.

TABLE III: Tukey HSD pairwise comparison of per-image F1-score between backbones.

Model 1	Model 2	p-value	Significant
inceptionv4	mobilenet_v2	0.7687	No
inceptionv4	resnet152	<0.0001	Yes
inceptionv4	vgg16	0.0001	Yes
mobilenet_v2	resnet152	<0.0001	Yes
mobilenet_v2	vgg16	0.0010	Yes
resnet152	vgg16	<0.0001	Yes

TABLE IV: Tukey HSD pairwise comparison of per-image IoU between backbones.

Model 1	Model 2	p-value	Significant
inceptionv4	mobilenet_v2	0.5074	No
inceptionv4	resnet152	<0.0001	Yes
inceptionv4	vgg16	0.0001	Yes
mobilenet_v2	resnet152	<0.0001	Yes
mobilenet_v2	vgg16	0.0010	Yes
resnet152	vgg16	<0.0001	Yes

Our findings highlight that lighter and more efficient architectures such as InceptionV4 and MobileNetV2 outperform deeper and more traditional networks like ResNet152 and VGG16, both in terms of segmentation performance and statistical significance. The Tukey HSD test confirmed these differences, reinforcing the hypothesis that backbone design plays a critical role in the effectiveness of U-Net models when operating under constrained data regimes. In particular, architectural elements such as depthwise separable convolutions and optimized feature reuse appear to contribute positively to model generalization and robustness.

While recent research has pointed to the use of simulators and digital twins as promising alternatives to overcome data scarcity and increase system adaptability in industrial settings, this work focused on the foundational step of model evaluation using real corrosion images. The integration of U-Net backbones within a simulated environment or digital twin remains an important avenue for future work. Additionally, the use of self-supervised pretraining and domain adaptation techniques could further enhance the performance of segmentation models under real-world deployment constraints.

References

[1] G. H. Koch, M. P. H. Brongers, N. G. Thompson, Y. P. Virmani, and J. H. Payer, "Corrosion cost and preventive strategies in the united states [final report]," United States. Federal Highway Administration, Tech. Rep. FHWA-RD-01-156, Mar. 2002.

[2] Association for Materials Protection and Performance, "Global cost of corrosion exceeds $2.5 trillion annually," 2021, accessed: 2025-06-09. [Online]. Available: https://tcdfillinois.org/news/the-high-cost-of-corrosion/

[3] M. Iannuzzi and G. S. Frankel, "Opportunities and challenges for corrosion science in the 21st century," *npj Materials Degradation*, vol. 6, no. 1, pp. 1–3, 2022. [Online]. Available: https://www.nature.com/articles/s41529-022-00224-w

[4] S. A. Bin Idris, F. A. Jafar, and N. Abdullah, "Study on corrosion features analysis for visual inspection & monitoring system: An ndt technique," *J. Teknol.*, vol. 77, no. 21, Dec. 2015.

[5] V. Vasagar, M. K. Hassan, A. M. Abdullah, A. V. Karre, B. Chen, K. Kim, N. Al-Qahtani, and T. Cai, "Non-destructive techniques for corrosion detection: A review," *Corros. Eng. Sci. Technol.*, vol. 59, no. 1, pp. 56–85, Feb. 2024.

[6] R. Senthilnathan, "Deep learning in vision-based automated inspection: Current state and future prospects," *Machine Learning in Industry*, pp. 159–175, 2021.

[7] J. E. See, C. G. Drury, A. Speed, A. Williams, and N. Khalandi, "The role of visual inspection in the 21st century," in *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1. SAGE Publications Sage CA: Los Angeles, CA, 2017, pp. 262–266.

[8] L. Leontaris, N. Dimitriou, A. Nikolousis, D. Tzovaras, and E. Papageorgiou, "Inspection of surface defects in metal processing industry using unet-based architectures," 2023.

[9] Z. Li and X. Liu, "Soldering defect segmentation method for pcb on improved unet," *Applied Sciences*, vol. 14, no. 16, p. 7370, 2024.

[10] I. Konovalenko, P. Maruschak, J. Brezinová, O. Prentkovskis, and J. Brezina, "Research of u-net-based cnn architectures for metal surface defect detection," *Machines*, vol. 10, no. 5, p. 327, 2022.

[11] S. K. Fondevik, A. Stahl, A. A. Transeth, and O. Ø. Knudsen, "Image segmentation of corrosion damages in industrial inspections," in *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2020, pp. 787–792.

[12] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, and A. K. Nandi, "Medical image segmentation using deep learning: A survey," *IET image processing*, vol. 16, no. 5, pp. 1243–1267, 2022.

[13] L. G. Barros, R. T. de Oliveira, J. P. M. Santana, and A. R. D. Ribeiro, "Inspection of surface defects in metal processing industry using unet-based architecture," in *Proceedings of the 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*. Rio de Janeiro, Brazil: IEEE, October 2023.

[14] I. Ahmed, K. Saleh, and A. Alzahrani, "Segmentation-based deep learning approach for surface defect detection in industrial products," in *2023 International Conference on Electrical and Computing Technologies and Applications (ICECTA)*. Ras Al Khaimah, UAE: IEEE, 2023.

[15] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *International journal of computer vision*, vol. 128, pp. 261–318, 2020.

[16] T. Van Erp, E. R. Da Silva, and M. Stenger, "Industrial digital twin as a facilitator for business model innovations in the manufacturing industry," *Procedia CIRP*, vol. 118, pp. 330–335, 2023.

[17] Y. Chen, Y. Ding, F. Zhao, E. Zhang, Z. Wu, and L. Shao, "Surface defect detection methods for industrial products: A review," *Applied Sciences*, vol. 11, no. 16, p. 7657, 2021.

[18] A. Saberironaghi, J. Ren, and M. El-Gindy, "Defect detection methods for industrial products using deep learning techniques: A review," *Algorithms*, vol. 16, no. 2, p. 95, 2023.

[19] R. Azad, E. K. Aghdam, A. Rauland, Y. Jia, A. H. Avval, A. Bozorgpour, S. Karimijafarbigloo, J. P. Cohen, E. Adeli, and D. Merhof, "Medical image segmentation review: The success of u-net," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[20] Z. Lv, H. Huang, L. Gao, J. A. Benediktsson, M. Zhao, and C. Shi, "Simple multiscale unet for change detection with heterogeneous remote sensing images," *IEEE Geoscience and Remote Sensing Letters*, vol. 19, pp. 1–5, 2022.

[21] P. Zhou, G. Zhou, Y. Li, Z. He, and Y. Liu, "A hybrid data-driven method for wire rope surface defect detection," *IEEE Sensors Journal*, vol. 20, no. 15, pp. 8297–8306, 2020.

[22] K. Shaikh, I. Hussain, and B. S. Chowdhry, "Wheel defect detection using a hybrid deep learning approach," *Sensors*, vol. 23, no. 14, p. 6248, 2023.

[23] sem, "cor dataset," https://universe.roboflow.com/sem-l2uic/cor-oumk1, jun 2025, visited on 2025-06-16. [Online]. Available: https://universe.roboflow.com/sem-l2uic/cor-oumk1

[24] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[25] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[27] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.

[28] C. Tan, F. Sun, T. Kong, W. Zhang, C. Yang, and C. Liu, "A survey on deep transfer learning," in *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27*. Springer, 2018, pp. 270–279.

[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of the ACM*, vol. 60, no. 6, pp. 84–90, 2017.

[30] A. Mazinanian, B. Kabiri, and V. Hashemi, "Research of u-net-based cnn architectures for metal surface defect detection," *Mathematics*, vol. 11, no. 17, p. 3693, 2023.

[31] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *2016 fourth international conference on 3D vision (3DV)*. Ieee, 2016, pp. 565–571.

[32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in python," *the Journal of machine Learning research*, vol. 12, pp. 2825–2830, 2011.

[33] M. Everingham, L. Van Gool, C. K. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International journal of computer vision*, vol. 88, pp. 303–338, 2010.

[34] J. W. Tukey, *Exploratory Data Analysis*. Reading, MA: Addison-Wesley, 1977.