

Comparative Analysis of Illumination Levels in the Performance of Neural Networks for Defect Detection in Metallic Surfaces

Alicia María Reyes-Duke¹; Pablo Ernesto Quiroz²

^{1,2}Universidad Tecnológica Centroamericana, (UNITEC), Honduras, aliciareyes@unitec.edu, quiroz25@unitec.edu

Abstract—Defect detection on metallic surfaces using artificial vision systems is highly sensitive to environmental disturbances such as reflections, shadows, and color temperature variations. These factors directly affect the performance of artificial neural networks (ANNs) during training and inference. This study evaluates the impact of three illumination ranges—optimal (150–300 Lux), moderate (350–500 Lux), and inadequate (550–1200 Lux)—on the performance of YOLO V5, YOLO V8, and Roboflow 3.0 for metallic surface defect detection. A dataset was collected from local metalworking workshops, and a spiral methodology was implemented to iteratively evaluate model behavior. Results show that optimal illumination conditions significantly improve performance, reaching a maximum mAP of 99.5% for iron surfaces and 88.7% for aluminum. The findings demonstrate that illumination control plays a critical role in ensuring reliable defect detection, particularly for highly reflective materials.

Keywords—Artificial vision, convolutional neural networks, data augmentation, illumination levels, metallic surface defects, YOLO.

I. INTRODUCTION

Industrial inspection processes increasingly rely on automated systems to improve reliability, reduce human error, and enhance operational efficiency. However, environmental factors such as illumination variability remain critical challenges in vision-based defect detection systems.

An example is the lack of standardization; by not planning the new technologies incorporated in the industry works it affects negatively the performance of industrial process [1].

The scope of technology nowadays is extensive; there is a tolerance for the integration of technological tools by adapting previous tools with the new ones or by using a complete autonomous system such as robots in different scales of work [2].

The metal mechanic industry is in constant development such as studies of tribology, production of components with complex geometry, defect detection and return prediction for elastic elements; [3] mention that there is a need to predict defects and results for a better performance in industrial process, by replacing manual tasks developed by humans to an automated system such as artificial vision which is one of the most popular alternatives nowadays[4].

Artificial vision is commonly used for supervision since this is a repetitive and exhausting task for human beings and can injury the sight sense by long time works. The most

common tools for artificial vision are neural networks. These are flexible tools and are an alternative for high repetitive processes, but the illumination levels are disturbances found in every environment and can lower the performance of these tools by using inappropriate levels.

The illumination systems are a fundamental part for every work, however, [5], mention that 30% to 50% of the accidents occurred during industrial work are caused by low illumination levels. [6-7] mention that illumination systems affect neural networks performance by the reflective properties of materials or by the generation of shadows by using low illumination levels.

An alternative for better clarity while developing a work under dark sceneries is to use artificial lightning such as hand spotlight or the cellphone flash, [8] mention that by using optimal illumination levels it can increase the performance of artificial vision tools in 15% independently if it's used artificial lightning.

The less estimated characteristic during ANNs studies are the illumination levels, [9] determine that illumination conditions affect its performance because these tools are based on the color analysis supplied by input image of the model being sensitive to the environment illumination.

The color temperature is a measure of illumination systems that indicates how warm or cold the light tone is, an example of warm temperatures are incandescent bulbs and cold temperatures are LED bulbs. [10] says that as the temperature color is changed the study color surface changes, affecting computerized vision systems by not having a stabilized color scene generating confusion in the model.

During the investigation of [11] was used both color temperatures for the designing of an ANN for defect detection over laminated metal sheets, the conclusion was that LED illumination source exposes defect characteristics better than incandescent does. During [12], [13] investigation it is observed a dedicated defect detection station using LED bulbs as light source which the results were around 96.5% and 98% of model precision.

The background is another characteristic that can generate confusion in the neural network models during training due to the texture of the background that can be like the study surface, nevertheless, if the texture of the background increases the contrast of the studied characteristic it can raise the performance [14].

A way to improve the performance of neural networks is by augmentations, which are a positive influence during

training which can be applied different filters, zoom on images, noise, blur, etc. During the investigation of [15] mosaic filter was used as a dataset augmentation and obtained a mAP of 94.7%.

During the investigation of [16] grayscale filter was used as a dataset augmentation, and mention that using this filter increases the positive detection of the model for the extraction of study characteristics.

In metal mechanic industry, there are some characteristics that need to be noticed instantly by supervision tasks, such as welding quality, corrosion and post machining defects. The three can be detected by human sight or artificial vision, but without the adequate illumination level this task could lead to miss important details which can evolve into greater defects reducing the lifetime of materials in a structure or machinery generating a resource loss.

The International Comision of Illumination (1997), International Organization for Standardization (2002) and Illuminating Engineering Society (2017) determined a standard illumination level depending on the area of work as shown in Table. 1.

TABLE I
ILLUMINATION LEVELS REQUIRED FOR DIFFERENT WORK AREAS

Work Area	Illumination Level [Lux]
Work Area	120-250
Assembly and Machinery	120-750
Electronic	200-2,000
Foundry	120-500

The illumination levels not only affect the performance of tasks, but it also increases the sight injuries in a work environment which it is estimated that 18% of the injuries are inadequate illumination levels reason and 1.5% of the workers finish with serious sight injuries [17].

During the recollection of images in a workshop could be present some disturbances generated by illumination sources, which can be:

Excessive illumination: The lightning is more pronounced and generates highlights which affect the image with imbalance of colors and overexposure, leading the neural network not being able to generalize patterns meaning that the model developed is confused.

Insufficient illumination: By having low lightning source it can generate shadows or clarity lack in images which some details can be missed when the model is developed.

Unstable illumination: Sequential variations like blinking lights in the environment can generate confusing results generating inconsistent data while analyzing the dataset, color temperature is another example that changes the perception of the extracted characteristics by changing the color surface in images.

Light direction: An alternative for the mentioned disturbances is using artificial light like hand spotlight or cellphone flash, however, this generates projection of shadows in the surface where the miss of details and texture confusion are the results.

Shadows: This is the most common disturbance in the context of light sources; shadows are the consequence of low illumination levels where the results of having presence of shadows in images are the incapacity of detecting characteristics and invisible part areas.

Vibrations: This disturbance is generated by machines operating close to the collecting data area or the movement of the component or surface that is studied generating blur images, change of position of the object and problems to generalize during training.

The advance of Artificial Intelligence tools in the last years has generated an impact on industrial processes, [18], [19] mention that it is possible to implement AI tools with conventional inspection tools like nondestructive tests and robotics systems with artificial vision to develop an autonomous tool to increase the performance of industrial processes and datasets nowadays can be generated by CADs to improve the quantity of studied characteristics[20].

Although several studies have explored metallic surface defect detection using convolutional neural networks, most investigations primarily focus on improving network architectures, preprocessing techniques, or dataset augmentation strategies. However, limited attention has been given to the quantitative evaluation of illumination levels during dataset acquisition and their influence on neural network performance.

Previous studies commonly evaluate defect detection under a single illumination condition or controlled laboratory environments, without comparing the robustness of multiple neural network architectures under different illumination ranges. Furthermore, metallic materials present highly reflective properties that generate glare, overexposure, and shadow-related disturbances, which directly affect feature extraction and model generalization.

Therefore, this study proposes a comparative analysis of different illumination ranges and their impact on the performance of Roboflow 3.0, YOLO V5, and YOLO V8 for metallic surface defect detection. The investigation also evaluates the influence of preprocessing and augmentation strategies to improve model robustness under unfavorable illumination conditions.

This investigation was conceived by the small number of studies related to the illumination levels as a variable of study and how it affects defect detection performance neural networks. Based on the related studies the structure for ANN training, it filters by case the images obtained depending on its illumination level, then the images pass through a preprocess to generate the dataset for training and finally analyze the results as shown in Fig. 1.

The proposed methodology includes controlled image acquisition under three illumination ranges, preprocessing techniques, augmentation strategies, and comparative ANN training to identify the illumination conditions that maximize detection performance while minimizing reflective disturbances.

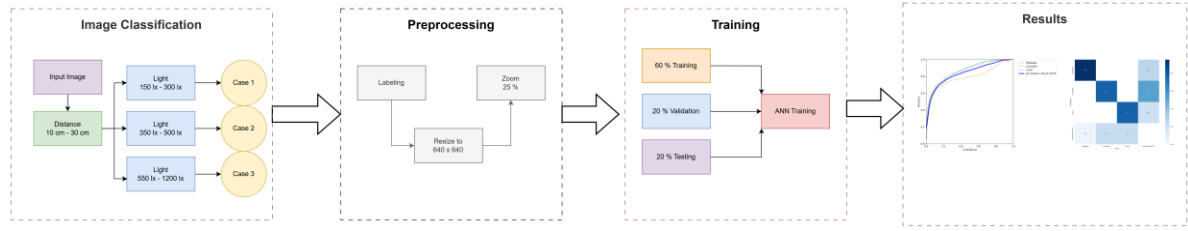


Fig. 1 Integrative image of investigation development.

II. METHODOLOGY

A spiral methodology was implemented to analyze risks and evaluate improvements during each development cycle. This methodology guarantees continuous feedback whenever new tools or techniques are introduced.

The spiral model consists of four phases: objective determination, risk analysis, development and testing, and planning for the next cycle. In this investigation, the primary objective was to achieve a mAP of 90% under two

illumination ranges. Risk analysis focused on factors such as inadequate color temperature and reflective disturbances that could reduce defect visibility during image acquisition.

During the development phase, data recollection and ANN training were performed using artificial vision tools, as shown in Fig. 2. Finally, the last phase consisted of defining improvements and objectives for the following experimental cycle.

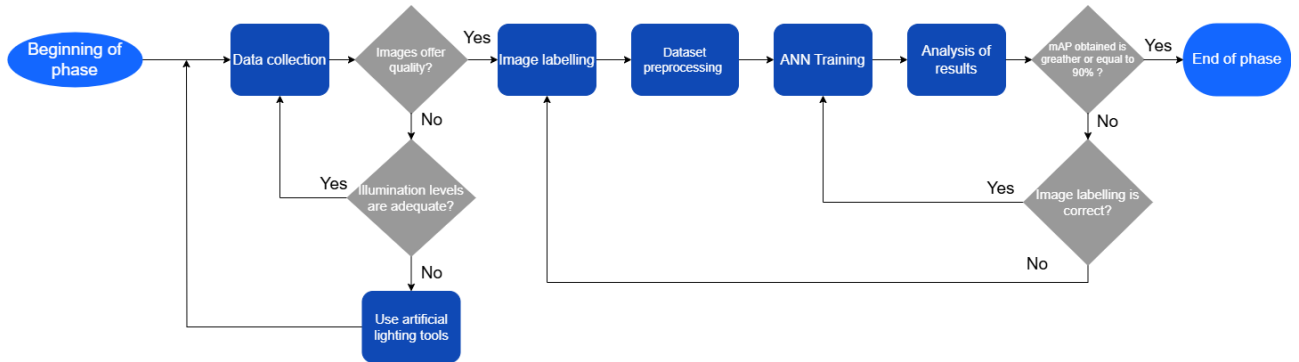


Fig. 2 Flow chart of the development and integration of artificial vision tools.

To generate a detailed dataset, photographic techniques such as plant and elevated views were implemented depending on the surface characteristics to analyze. Bounding box and polygon segmentation annotation techniques were used according to the geometric complexity of each defect.

Illumination levels were measured using Lux for Android and Luxmeter LM-3000 for iPhone, both based on the environmental light sensor integrated into the devices. Images were captured using an iPhone XS Max with a 12 MP camera and converted from HEIF to JPG/JPEG format for compatibility. Photographs were taken at an approximate distance of 30 cm from the metallic surface to maximize detail visibility.

Roboflow and YOLO were implemented as artificial vision tools. Roboflow 3.0 was used for image labeling, augmentation, resizing, and training processes, while YOLO V5 and YOLO V8 were used exclusively for model training and performance evaluation. *A. Experimental Setup*

The experimental setup was designed to evaluate the performance of different artificial neural network architectures under controlled illumination conditions. Metallic surfaces

containing defects such as corrosion, cracks, patches, and pitted regions were photographed under three illumination ranges: optimal illumination (150–300 Lux), moderate illumination (350–500 Lux), and inadequate illumination (550–1200 Lux).

Images were captured using an iPhone XS Max camera with a resolution of 12 megapixels at an approximate distance of 30 cm from the metallic surface to maximize defect visibility and reduce background interference. Cold LED illumination was selected due to its ability to expose surface details more effectively compared to warm illumination sources.

The dataset was divided into training, validation, and testing subsets using a 60%, 20%, and 20% distribution, respectively.

Augmentation procedures were applied exclusively to the training subset after dataset partitioning to prevent data leakage between training and testing stages.

The dataset partitioning process was performed before augmentation procedures to prevent duplicated or transformed samples from appearing across training and testing subsets.

TABLE II
EXPERIMENTAL PARAMETERS

Parameter	Value
Dataset Split	60% Training, 20% Validation, 20% Testing
Training Epochs	150
Dataset Sizes Evaluated	100, 200, and 300 images
Illumination Ranges	150–300 Lux, 350–500 Lux, 550–1200 Lux
ANN Models	YOLO V5, YOLO V8, Roboflow 3.0
Image Capture Distance	30 cm
Illumination Source	Cold LED
Annotation Techniques	Bounding Box, Polygon Segmentation
Augmentation Techniques	Grayscale, Zoom
Evaluation Metric	mAP
Defect Types	Cracks, Corrosion, Patches, Pitted Surfaces

The dataset was partitioned into 60% training, 20% validation, and 20% testing subsets. All models were trained for 150 epochs under the three defined illumination ranges to evaluate convergence behavior and detection robustness under different environmental lighting conditions.

III. RESULTS

The performance of the evaluated artificial neural network (ANN) models was analyzed under three different illumination ranges: optimal (150–300 Lux), moderate (350–500 Lux), and inadequate (550–1200 Lux). The evaluation metric used was mean Average Precision (mAP), measured across training epochs. In all performance graphs, the horizontal axis represents the number of training epochs, and the vertical axis represents the mean Average Precision (mAP).

A. Performance under Optimal Illumination (150–300 Lux)

Under optimal illumination conditions, all evaluated models achieved stable performance and high convergence rates. As shown in Fig. 3, YOLO V8 reached the highest mAP of 99.5% before the 60th epoch, demonstrating faster convergence than YOLO V5 and Roboflow 3.0. Although the remaining models achieved similar mAP values, they required a greater number of epochs. These results confirm that controlled illumination conditions improve defect detection and model generalization.

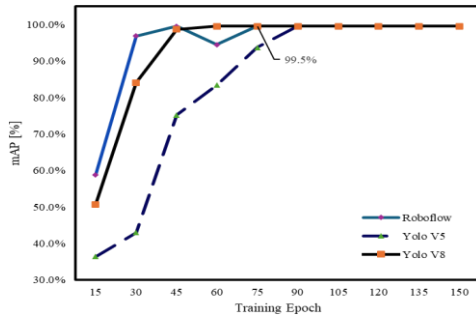


Fig. 3 Artificial neural networks models performance through epochs for 150–300 Lux.

B. Performance under Moderate Illumination (350–500 Lux)

When trained with datasets collected under moderate illumination levels, performance decreased relative to optimal conditions. As presented in Fig. 4, YOLO V8 achieved a maximum mAP of 92.2% at approximately 120 epochs.

However, YOLO V5 and Roboflow 3.0 remained below 90% mAP.

The results suggest that moderate illumination introduces additional variability, requiring longer training time and affecting model stability.

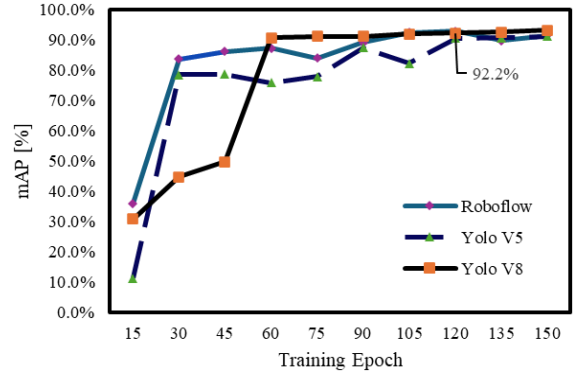


Fig. 4 Artificial neural networks models performance through epochs for 350–500 Lux.

C. Performance under Inadequate Illumination (550–1200 Lux)

Significant performance degradation was observed when models were trained using datasets captured under excessive illumination levels. As shown in Fig. 5, none of the evaluated models reached 90% mAP. The highest performance was achieved by YOLO V5 with 68.8% mAP, while YOLO V8 and Roboflow 3.0 remained below 65%.

This decline confirms that excessive illumination levels generate reflections and overexposed regions that reduce feature discriminability.

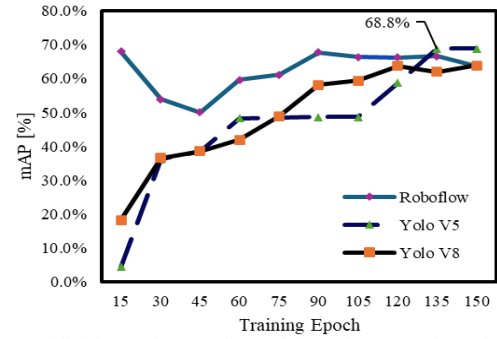


Fig. 5 Artificial neural networks models performance through epochs for 550–1200 Lux.

Table III summarizes the performance degradation experienced by all evaluated models as illumination levels increased. YOLO V8 achieved the highest performance under optimal and moderate illumination conditions, while YOLO V5 demonstrated slightly better robustness under excessive illumination levels.

TABLE III
PERFORMANCE COMPARISON UNDER DIFFERENT ILLUMINATION LEVELS

Illumination Range [Lux]	YOLOV5 mAP [%]	YOLOV8 mAP [%]	Roboflow3.0 mAP [%]
150–300	98.7	99.5	97.9
350–500	88.1	92.2	87.4
550–1200	68.8	64.5	62.3

D. Effect of Data Augmentation

The implementation of augmentation techniques positively influenced model performance. Fig. 6 shows that without augmentations, models under moderate and inadequate illumination conditions achieved mAP values below 70%. After applying zoom and grayscale augmentations, performance improved significantly.

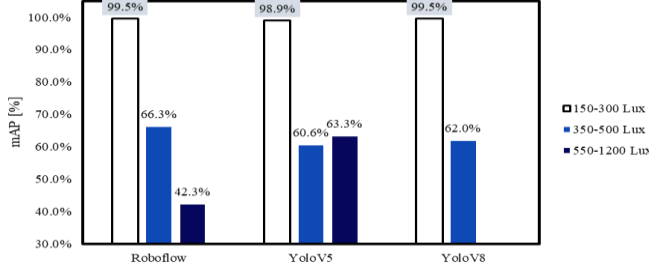


Fig. 6 Artificial Neural Networks performance comparison without augmentations.

As shown in Fig. 7, moderate illumination levels reached mAP values above 90%, while inadequate illumination levels improved to approximately 70%.

TABLE IV

EFFECT OF DATA AUGMENTATION ON MODEL PERFORMANCE

Illumination Condition	Without Augmentation mAP [%]	With Augmentation mAP [%]
Moderate	<70	>90
Inadequate	<65	~70

The implementation of augmentation techniques significantly improved model generalization under unfavorable illumination conditions shown in Table IV.

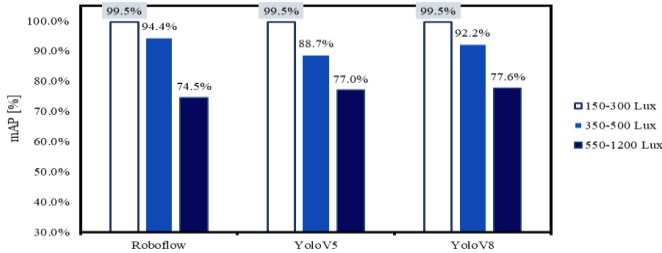


Fig.7 Artificial Neural Networks performance comparison with augmentations.

E. Effect of Dataset Quantity

The relationship between dataset size and model robustness was evaluated, as shown in Fig. 8. Increasing dataset quantity under optimal illumination resulted in minor performance variation (approximately 5%). However, under moderate illumination the variation increased to approximately 10%, and under inadequate illumination, performance degradation reached approximately 30%.

These results demonstrate that dataset expansion alone cannot compensate for unfavorable illumination conditions.

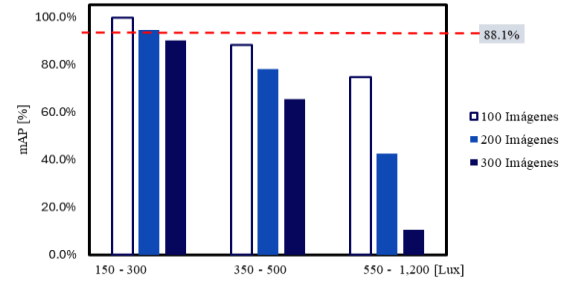


Fig. 8 Artificial Neural Networks model performance with different dataset quantity.

IV. DISCUSSION

The experimental results demonstrated that illumination conditions significantly influence defect detection performance on metallic surfaces. Under optimal illumination conditions (150–300 Lux), all evaluated models achieved high mAP values, indicating that controlled lighting improves feature visibility and reduces reflective disturbances.

As illumination levels increased, detection performance decreased due to glare and overexposed regions commonly generated on metallic materials. These disturbances reduce contrast between defects and the surrounding surface, limiting the ability of convolutional neural networks to extract stable features.

The comparative analysis showed that YOLO V8 achieved the best overall performance and faster convergence under favorable illumination conditions. However, under excessive illumination levels, performance differences between YOLO V8, YOLO V5, and Roboflow 3.0 became less significant, suggesting that environmental disturbances can reduce the advantages of more advanced network architectures.

Data augmentation strategies improved model robustness under moderate illumination conditions by increasing variability during training. Nevertheless, augmentation alone was insufficient to compensate for severe reflective artifacts. Similarly, increasing dataset size improved training stability but did not eliminate performance degradation under unfavorable illumination conditions.

Overall, the findings demonstrate that reliable metallic surface defect detection requires appropriate illumination control, optimized preprocessing techniques, and robust neural network architectures.

V. LIMITATIONS

This investigation was limited to controlled workshop environments and three predefined illumination ranges. Additional environmental disturbances such as dust, dynamic shadows, vibrations, and motion blur were not experimentally evaluated. Furthermore, only three ANN architectures were considered during the comparative analysis. Future investigations should explore adaptive illumination normalization techniques, larger datasets, and real-time industrial deployment scenarios.

VI. CONCLUSIONS

This study evaluated the influence of illumination levels on the performance of convolutional neural network models for metallic surface defect detection. The results demonstrated that illumination conditions significantly affect detection accuracy and model convergence behavior.

Under optimal illumination conditions (150–300 Lux), all evaluated models achieved high mAP values, with YOLO V8 presenting the best overall performance and fastest convergence. In contrast, excessive illumination levels generated reflections and overexposed regions that reduced detection accuracy across all models.

Data augmentation techniques improved model robustness under moderate illumination conditions; however, they were insufficient to fully compensate for severe reflective disturbances. Additionally, increasing dataset size improved training stability but did not eliminate the negative impact of unfavorable illumination conditions.

Overall, the findings confirm that reliable metallic surface defect detection requires appropriate illumination control, optimized preprocessing strategies, and robust deep learning architectures. Future work should focus on adaptive illumination correction and real-time industrial deployment.

REFERENCES

- [1] C. Caux, F. David, y H. Pierreval, "Implementation of delayed differentiation in batch process industries: a standardization problem", *Int. J. Prod. Res.*, vol. 44, núm. 16, pp. 3243–3255, ago. 2006, doi: 10.1080/002075405000521543.
- [2] E. O. Serrano-López, R. A. Palomeque-Castellanos, H. A. Moreno, y J. L. Ordóñez-Avila, "Design of a Modular Pipe Inspection Robot with Internal Damage Detection Using a Convolutional Neural Network", en 2023 IEEE Central America and Panama Student Conference (CONESCAPAN), sep. 2023, pp. 151–156. doi: 10.1109/CONESCAPAN60431.2023.10328423.
- [3] C.-C. Grigoraş, Ştefan Coşa, y V. Zichil, "INTEGRATING NEURAL NETWORKS INTO SHEET METAL FORMING: A REVIEW OF RECENT ADVANCES AND APPLICATIONS", *J. Eng. Stud. Res.*, vol. 30, núm. 1, Art. núm. 1, jul. 2024, doi: 10.29081/jesr.v30i1.005.
- [4] D. F. N. Rivera y A. M. R. Duke, "Work Safety Assessment through Contextual Analysis with Computer Vision", en 2020 5th International Conference on Control and Robotics Engineering (ICCRE), Osaka, Japan: IEEE, abr. 2020, pp. 207–212. doi: 10.1109/ICCRE49379.2020.9096264.
- [5] Y. Serikov, L. Nazarenko, y K. Serikova, "Non-Visual Exposure to Light as a Production Factor of the Influence of Lighting of the Working Area on Labor Productivity and Safety of Workers", *Metrol. Instrum.*, pp. 35–39, ene. 2020, doi: 10.33955/2307-2180(6)2019.35-39.
- [6] C. Hu et al., "Impact of Light and Shadow on Robustness of Deep Neural Networks", el 23 de mayo de 2023, arXiv: arXiv:2305.14165. doi: 10.48550/arXiv.2305.14165.
- [7] R. Vengaloor y R. Muralidhar, "Deep Learning Based Feature Discriminability Boosted Concurrent Metal Surface Defect Detection System Using YOLOv-5s-FRN", *Int. Arab J. Inf. Technol.*, vol. 21, núm. 1, 2024, doi: 10.34028/iajit/21/1/9.
- [8] X. Zhang, N. Jia, y I. Ivrisstizts, "A study of the effect of the illumination model on the generation of synthetic training datasets", *ArXiv*, jun. 2020, Consultado: el 5 de febrero de 2025. [En línea]. Disponible en: <https://www.semanticscholar.org/paper/fa6f4b16679b6fe3a7aa133ab0a0786e24409eb0>
- [9] A. Kolaman, D. Malowany, R. R. Hagege, y H. Guterman, "Light Invariant Video Imaging for Improved Performance of Convolution Neural Networks", *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, núm. 6, pp. 1584–1594, jun. 2019, doi: 10.1109/TCSVT.2018.2846695.
- [10] V. C. Cardei, B. Funt, y K. Barnard, "Estimating the scene illumination chromaticity by using a neural network", *JOSA A*, vol. 19, núm. 12, pp. 2374–2386, dic. 2002, doi: 10.1364/JOSAA.19.002374.
- [11] P. Maruschak et al., "Surface Illumination as a Factor Influencing the Efficacy of Defect Recognition on a Rolled Metal Surface Using a Deep Neural Network", *Appl. Sci.*, vol. 14, núm. 6, Art. núm. 6, ene. 2024, doi: 10.3390/app14062591.
- [12] F. Çerezci et al., "Online Metallic Surface Defect Detection Using Deep LearningOnline Metallic Surface Defect Detection Using Deep Learning", el 16 de julio de 2020, In Review. doi: 10.21203/rs.3.rs-41274/v1.
- [13] L. Song, X. Li, Y. Yang, X. Zhu, Q. Guo, y H. Yang, "Detection of Micro-Defects on Metal Screw Surfaces Based on Deep Convolutional Neural Networks", *Sensors*, vol. 18, núm. 11, Art. núm. 11, nov. 2018, doi: 10.3390/s18113709.
- [14] C. Hu, J. Yao, W. Wu, W. Qiu, y L. Zhu, "A Lightweight Reconstruction Network for Surface Defect Inspection", el 25 de diciembre de 2022, arXiv: arXiv:2212.12878. doi: 10.48550/arXiv.2212.12878.
- [15] C. Wang, Z. Zhou, y Z. Chen, "An Enhanced YOLOv4 Model With Self-Dependent Attentive Fusion and Component Randomized Mosaic Augmentation for Metal Surface Defect Detection", *IEEE Access*, vol. 10, pp. 97758–97766, 2022, doi: 10.1109/ACCESS.2022.3203198.
- [16] O. Essid, H. Laga, y C. Samir, "Automatic detection and classification of manufacturing defects in metal boxes using deep neural networks", *PLOS ONE*, vol. 13, núm. 11, p. e0203192, nov. 2018, doi: 10.1371/journal.pone.0203192.
- [17] J. L. Ordoñez-Avila, E. R. T. Maldonado, and I. Magomedov, "Water Generation Based on Condensation Controlled by Gray Scale and Artificial Vision," *ICCT 2022*, DOI: 10.1109/ICCT56057.2022.9976774.
- [18] D. Merayo, A. Rodríguez-Prieto, y A. M. Camacho, "Prediction of Mechanical Properties by Artificial Neural Networks to Characterize the Plastic Behavior of Aluminum Alloys", *Materials*, vol. 13, núm. 22, Art. núm. 22, ene. 2020, doi: 10.3390/ma13225227.
- [19] D. J. Atha y M. R. Jahanshahi, "Evaluation of deep learning approaches based on convolutional neural networks for corrosion detection", *Struct. Health Monit.*, vol. 17, núm. 5, pp. 1110–1128, sep. 2018, doi: 10.1177/1475921717737051.
- [20] M. Ángel Castro-Rivera y A. María Reyes-Duke, "Towards a Smarter CAD: Application of the Integration of Artificial Vision and Neural Networks in Aided Design", en 2023 IEEE 41st Central America and Panama Convention (CONCAPAN XLI), Tegucigalpa, Honduras: IEEE, nov. 2023, pp. 1–6. doi: 10.1109/CONCAPANXLI59599.2023.10517573.