Revista de Informática Teórica e Aplicada - RITA - ISSN 2175-2745
Vol. 32, Num. 1 (2025) 40-46

RESEARCH ARTICLE

Composite Material Defect Segmentation Using Deep Learning Models and Infrared Thermography

Segmentação de Defeitos em Materiais Compósitos Usando Redes Neurais Profundas e Termografia Infravermelha

Iago Garcia Vargas[1]*, Henrique Fernandes[2,1]

Abstract: For non-destructive assessment, the segmentation of infrared thermographic images of carbon fiber composites is a critical task in material characterization and quality assessment. This study focuses on applying image processing techniques, particularly adaptive thresholding, alongside neural network models such as U-Net and DeepLabv3 for infrared image segmentation tasks. An experimental analysis was conducted on these networks to compare their performance in segmenting artificial defects from infrared images of a carbon-fibre reinforced polymer sample. The performance of these models was evaluated based on the F1-Score and Intersection over Union (IoU) metrics. The findings reveal that DeepLabv3 demonstrates superior results and efficiency in segmenting patterns of infrared images, achieving an F1-Score of 0.94 and an IoU of 0.74, showcasing its potential for advanced material analysis and quality control.
Keywords: Infrared Thermography — Segmentation — Deep Learning — Composite Materials

Resumo: Na avaliação não destrutiva, a segmentação de imagens geradas por termográfia infravermelho de compósitos de fibra de carbono é uma tarefa crítica, tanto na caracterização do material quanto na avaliação de sua qualidade. Este estudo foca na aplicação de técnicas de processamento de imagem, particularmente threshold adaptativo, juntamente com modelos de redes neurais como U-Net e DeepLabv3 para segmentação de imagens infravermelhas. Foi realizada uma análise experimental dessas redes para comparar seu desempenho na segmentação de defeitos artificiais em imagens infravermelhas de uma amostra de polímero reforçado com fibra de carbono. O desempenho desses modelos foi avaliado com base nas métricas de F1-Score e Intersection over Union (IoU). Os resultados revelam que a rede DeepLabv3 tem resultados superiores e maior eficiência na segmentação de padrões de imagens infravermelhas, alcançando um F1-Score de 0,94 e um IoU de 0,74, destacando seu potencial para análise de materiais compósitos e no controle de qualidade.
Palavras-Chave: Termografia Infravermelha — Segmentação — Aprendizado Profundo — Materiais Compósitos

[1] *Falculty of Computing, Federal University of Uberlândia (UFU), Brazil*
[2] *IVHM Centre, Faculty of Engineering and Applied Sciences, Cranfield University (CU), United Kingdom*
***Corresponding author**: iagogarcia@ufu.br
DOI: http://dx.doi.org/10.22456/2175-2745.143066 • **Received:** 09/10/2024 • **Accepted:** 02/12/2024
CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introduction

Composite materials are important structural materials that have been extensively utilized in fields like aerospace and wind turbine blade construction [1]. Defects that occur during the manufacturing and handling of laminated composites can present potential safety risks to devices and may lead to incalculable losses [2].

The advent of composites as a unique material class happened in the mid-20th century, marked by the production of intentionally designed multiphase composites, such as Carbon-Fibre-Reinforced Polymers (CFRP) [3]. Advanced image processing techniques, such as adaptive thresholding, along with deep learning approaches such as deep neural networks, are increasingly being applied to segment thermographic images of CFRP to accurately identify structural defects [4]. Given the critical role of CFRP in industries like aerospace and automotive, ensuring the material's integrity is extremely important [5]. Defects such as cracks, delaminations, or voids can significantly impair the material's performance and durability. Thus, the implementation of sophisticated Non-Destructive Testing (NDT) methods is vital to detect and quantify these flaws without causing damage to the material [6].

In this context, thermographic imaging offers a non-invasive and effective way to assess the condition of CFRP components. However, raw thermographic data often contains noise and requires advanced processing techniques to accurately identify defects [7]. In this context, image processing techniques, as well as deep learning techniques, are commonly employed where the challenge is to differentiate relevant thermal patterns differentiating defects from background noise [4].

This study introduces a comprehensive approach that integrates infrared thermography and deep learning techniques to enhance the accuracy of defect segmentation in infrared images of sample. We utilize metrics such as the F1-Score and Intersection over Union (IoU) to evaluate our models' performance and to validate the effectiveness of our proposed models in segmenting defects. The paper details the development of deep learning models, emphasizing the adaptations and optimizations made for thermographic data, and compares the performance of these algorithms for defect segmentation and identification. Significantly, our approach leverages a curated dataset of thermographic images, annotated with expert knowledge, to train and test the deep learning models. This dataset covers three types of defects with different dimensions (2x2, 3x3, 4x4mm) at different depths, providing a comprehensive basis for assessing the effectiveness of the models.

The paper is organized as follows: Section 2 delves into the materials and methods, explaining the image processing techniques and the rationale behind their selection, as well as the deep learning techniques used. This section also describes the dataset preparation and the training process for the deep learning models. Section 3 presents the results of image segmentation and defect detection, comparing the effectiveness of the employed methods. Section 4 concludes the study with a discussion of the findings and suggestions for future research.

2. Materials and Methods

2.1 Active thermography

Active infrared thermography is a NDT technique that utilizes the principles of thermodynamics and infrared imaging to analyze the thermal properties of materials. In this method, a heat source is applied to the surface of the material under inspection, causing a localized temperature increase. The material's thermal response is then recorded using an infrared camera, which captures the emitted infrared radiation. By analyzing the thermal patterns captured by the camera, valuable information about the material's internal structure, defects, and anomalies can be inferred [1].

Pulsed thermography (PT) is a variant of active infrared thermography wherein the heat source is applied in short pulses instead of continuously. This method allows for better control over the heating process and can enhance the detection sensitivity, especially for small or shallow defects. PT offers advantages in terms of speed, efficiency, and versatility, making it a valuable tool in various fields such as aerospace, automotive, and materials science for quality control and defect detection purposes.

2.2 Adaptive Threshold

Adaptive thresholding is an image processing technique that adapts to local intensity variations within an image to perform segmentation [8]. This technique is suitable for thermographic analyses of infrared images, whose complex thermal properties and structural heterogeneity can result in images with localized and variable thermal contrasts [9].

In active infrared thermography inspection, the heat propagation through the material can be altered by internal discontinuities such as delaminations, porosity, or inclusions, which modify the local thermal conductivity and, consequently, the observed heat distribution [10]. This heat distribution can be seen on the surface of the material using an infrared camera. Temperature differences and contrast is seen as pixel intensity variation on the infrared images. The Adaptive thresholding method allows for the isolation of these contrasting anomalies by adjusting the segmentation thresholds according to local mean or intensity variations.

Formally, adaptive thresholding is defined as follows in Equation (1):

For a pixel $p(x,y)$ in a infrared image I, the adaptive threshold $T(x,y)$ can be calculated as:

$$T(x,y) = m(x,y) \times \left(1 + k \times \left(\frac{s(x,y)}{R} - 1\right)\right) \tag{1}$$

where $m(x,y)$ is the local mean intensity, which is calculated by averaging the pixel values in a predefined neighborhood around the pixel $p(x,y)$. This neighborhood consists of a small window of pixels centered at (x,y). The local mean intensity is used to adapt the thresholding process to local variations in intensity across the image. The $s(x,y)$ is the local standard deviation, k is an empirical adjustment parameter, and R is the dynamic range of pixel intensities.

To apply adaptive thresholding for segmentation, we first calculate the threshold $T(x,y)$ for each pixel $p(x,y)$ using the formula provided in Equation (1). Once the threshold is determined, the segmentation process involves comparing the intensity of each pixel in the image $I(x,y)$ to its corresponding threshold $T(x,y)$. If the pixel intensity $I(x,y)$ exceeds the threshold $T(x,y)$, the pixel is classified as part of a defect; otherwise, it is classified as part of the background. This pixel-wise classification creates a binary segmentation mask where pixels belonging to defects are labeled as 1 and the background as 0. This mask is then used to isolate and analyze the defect regions in the infrared images.

2.3 U-Net

The U-Net neural network is an architecture widely used for image segmentation, being particularly prominent in medical and industrial applications [11]. Initially developed by [12] in 2015, the U-Net is notable for its efficiency in working with a limited number of training samples and for the capacity to capture contextual and localization details simultaneously.

The U-Net architecture is characterized by its symmetric shape as shown in Figure 1. It consists of a contracting path (encoder) and an expansive path (decoder). The contracting path reduces the image dimensions while increasing the depth of the features. The expansive path uses up-sampling operations and convolutions to reconstruct the image from the encoded features, increasing the resolution and the specificity of the location.

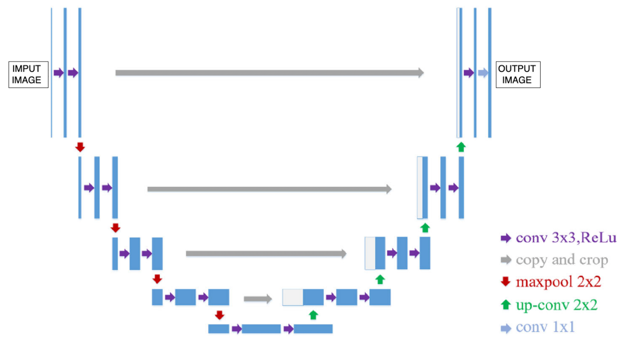


Figure 1. The architecture of the U-Net (adapted from [13]).

The model was implemented using the PyTorch library, leveraging specific GPU processing capabilities, which are crucial for efficient training with large image datasets [14]. The model's consists of repeated blocks of convolutions and activation operations. Each block executes two sequences of operations, including a 2D convolution, batch normalization, and the ReLU activation function, applied sequentially.

The contraction path follows a chained structure, where each block is followed by a max pooling operation for dimensionality reduction. This path consists of four levels of double convolutions with a progressive increase in the number of filters, starting from 64 up to 512 filters at the last level. The expansion path uses bilinear upsampling operations to increase the spatial resolution of the learned features. After each upsampling, a concatenation with the corresponding feature map from the contraction path is performed, followed by another convolution block. After the completion of the contraction and expansion paths, a final convolution is applied to map the resulting features into a binary segmentation map output, using a convolutional layer with one filter.

The model is optimized using the Adam algorithm with an initial learning rate of 1×10^{-2}. Logistic loss with logits (BCEWithLogitsLoss) was used. Training is conducted over 15 epochs, adjusting the parameters in a parallel processing environment on the GPU. The GPU used in this training was an NVIDIA Tesla T4, offering 16 GB of memory, 2,560 CUDA cores, and 320 Tensor Cores. It delivers up to 8.1 TFLOPS (FP32) and 130 TOPS (INT8).

2.4 DeepLabv3

DeepLabv3 is a deep neural network architecture developed for semantic image segmentation tasks, which has proven effective in identifying and segmenting complex features across various applications. DeepLabv3 incorporates "atrous convolution modules" (known as dilated convolutions) that enable capture of contextual information without loss of spatial resolution [15]. In the application for defect segmentation in infrared images of CFRP, DeepLabv3 is advantageous due to its ability to handle the variations in texture and temperature that characterize infrared images [16]. Additionally, its structure allows for the incorporation of a spatial classification module, known as "Atrous Spatial Pyramid Pooling" (ASPP), which uses various dilation rates to capture context at multiple scales [15].

Figure 2 presents the DeepLabv3 architecture. It begins with convolutional layers that progressively reduce the image resolution to extract key features. To capture a wider context without increasing computational complexity, atrous (or dilated) convolutions are applied, effectively expanding the field of view. The central element, the Atrous Spatial Pyramid Pooling (ASPP) module, uses atrous convolutions at different rates, along with global pooling, to capture detailed features across multiple scales.

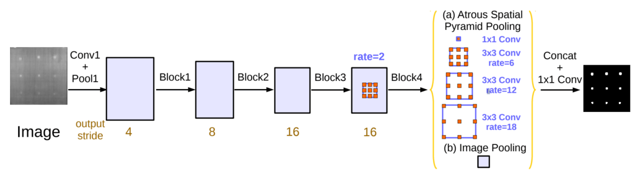


Figure 2. The architecture of DeepLabv3 (adapted from [15]).

The DeepLabv3 implementation initializes with a backbone networks, setup to 'resnet101' due to its robust feature extraction capabilities [17]. This choice benefits from pre-training on a vast dataset, enhancing the model's ability to generalize from diverse visual representations.

The developed model adapts the first convolutional layer to adjust the specified number of input channels - in our case 3-channel RGB. The output of the model is adjust to the specific needs of the task by modifying the final convolutional layer in the classifier module. This modification ensures the output channels match the binary segmentation.

The model is employed in a computational setup optimized for performance, using the same GPU as the U-Net setup. Training is conducted using the Binary Cross-Entropy with Logits Loss (BCEWithLogitsLoss), which is suitable for binary classification tasks. The Adam optimizer is used for learning the model parameters with an initial learning rate of 1×10^{-4}, and the training process spans 15 epochs.

2.5 CFRP Dataset

The dataset used in this study is well explained by [18]. Briefly, a unidirectional carbon/PEEK laminate with a fiber volume fraction of 61% was inspected using different thermographic methods. The laminate, measuring 100×100 mm, incorporated artificial Kapton tape inserts of varying sizes (2x2, 3x3, 4x4mm) placed at specific layers before molding, as show in Figure 3. For this study, only the PT inspections were used, where a midwave infrared (MWIR) camera captured thermal profiles following short, intense pulse. The infrared camera used acquired 640 x 512 pixel images at a frame rate of 55 Hz and the experiment generated 1053 images which form the dataset used in this study. For each pixel, for our models, it could either belongs to a defect or a sound (background) area. Each pixel, of each images, was also labeled as defect or sound area.

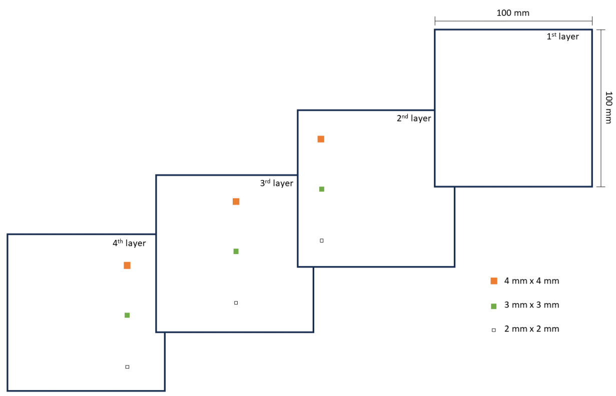


Figure 3. Position of the defects on the layers of the laminate (adapted from [18]).

2.6 Training, Validation, and Test Data

The dataset comprises 1053 images of CFRP samples, which were divided into training, validation, and test sets. Specifically, 70% of the images (737 images) were used for training, 15% (158 images) for validation, and the remaining 15% (158 images) for testing. This split ensures a balanced representation of the defects in each subset, facilitating robust model training and evaluation. Prior to training, each image was normalized to enhance contrast and detail.

2.7 Evaluation Metrics

For this study, F1-Score and IoU were used as evaluation metrics. The F1-Score is a measure of a test's or model's accuracy, considering both precision and recall. It is particularly useful when the classes are imbalanced. The F1-Score is the harmonic mean of precision and recall, providing a balance between these two metrics [19]. It is calculated using the Equation (2):

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{2}$$

IoU is a common metric for evaluating the performance of segmentation tasks. The metric quantifies how much a model's prediction overlaps with the true object area [20]. IoU is calculated using Equation (3), by dividing the intersection area between the predicted bounding box and the actual bounding box by the size of the union of the two boxes.

$$IoU = \frac{\text{Intersection Area}}{\text{Union Area}} \tag{3}$$

The closer the IoU is to 1, the more accurate the model's detection. An IoU close to 0 indicates low accuracy in predicting object locations.

3. Results and Discussion

In this study, a PT experiment was conducted to investigate the effectiveness of different image processing techniques for defect segmentation of infrared images of a CFRP. To segment the defects, three approaches were employed: adaptive thresholding, used as a baseline, U-Net model and DeepLabv3 model. The results of the segmentation applied to one of the images are presented in Figure 4. The image shows 6 defects (2x2, 3x3, 4x4mm). Figure 4(a) shows the original, Figure 4(b) the image annotated manually by an expert, Figure 4(c) shows the mask for segmentation, Figure 4(d) shows the segmentation using adaptive threshold, Figure 4(e) shows the segmentation using U-Net, and Figure 4(f) shows the segmentation using DeepLabv3.

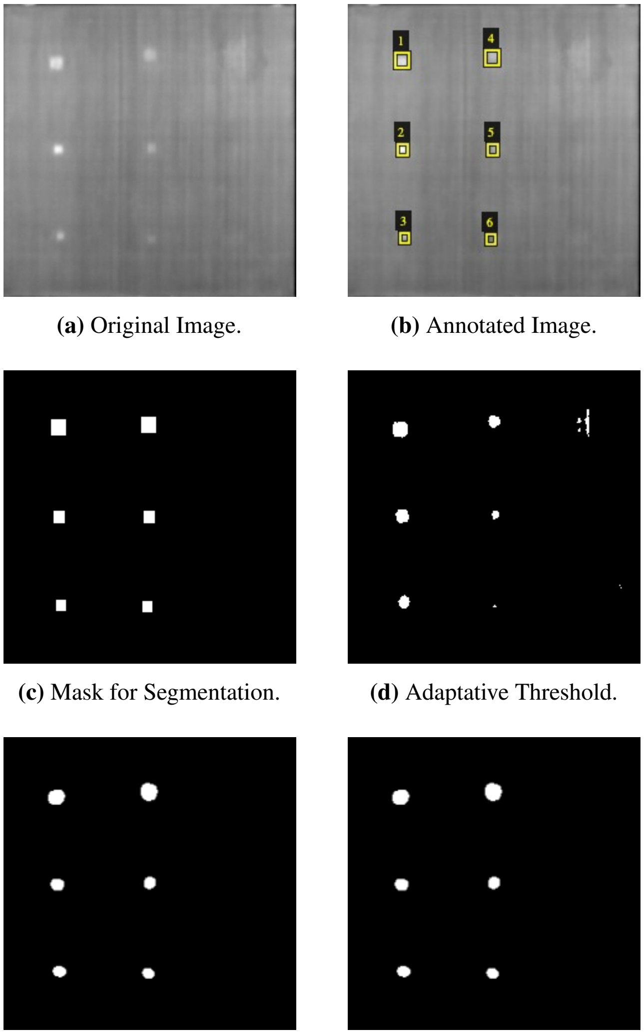
The mask for segmentation shown in Figure 4(c) is a binary mask created from manual annotations of the defects. Each pixel in the mask is labeled as either a defect (value 1) or non-defect (value 0). This mask serves as a ground truth reference for evaluating the segmentation performance of different methods. The manual annotation process involves marking the exact locations and boundaries of defects in the infrared images, ensuring that the segmentation models have accurate targets for training and validation.

Initially, an adaptive thresholding technique was applied to segment the infrared images. This approach yielded an average F1-Score of 0.7519 and an IoU metric of 0.2375 on the test dataset. Despite its simple and quick implementation, the results show that this technique had limited performance in terms of segmentation accuracy.

Subsequently, a U-Net neural network was trained to provide a more robust segmentation model. The results were much better when compared to the first one, achieving an IoU of 0.6141 and an F1-Score of 0.9142 on the test dataset. The improvement in performance compared to adaptive thresholding suggests that a deep learning based approach is more effective in capturing the complexities associated with thermal phenomena present in the PT experiment.

Finally, the DeepLabv3 neural network, an architecture for semantic segmentation, was applied. This technique produced the best results, with an F1-Score of 0.9406 and an IoU of 0.7419 on the test dataset. The high performance of DeepLabv3 indicates a superior ability in precisely segmenting areas of interest in complex thermographic images.

Both the U-Net and DeepLabv3 models were trained using the Adam optimization algorithm, with an initial learning rate set to 1×10^{-2}. The Binary Cross-Entropy with Logits



(a) Original Image. **(b)** Annotated Image.

(c) Mask for Segmentation. **(d)** Adaptative Threshold.

(e) U-Net. **(f)** DeepLabv3.

Figure 4. Comparison of the different methods applied.

Loss (`BCEWithLogitsLoss`) function was employed to compute the loss, combining the benefits of sigmoid activation and cross-entropy for better stability. The training process was carried out over 60 epochs, with a mini-batch size of 16.

To accelerate computations and ensure efficient learning, the models were trained in a parallel processing environment leveraging GPU hardware. During training, the learning rate was adjusted using a learning rate scheduler to prevent overfitting and accelerate convergence. All experiments were implemented in TensorFlow, utilizing an NVIDIA T4 GPU with 16 GB of VRAM in the hardware environment.

Figure 5 presents a summary of the results of the three tested approaches. The IoU achieved in this study may be linked to the thermal diffusion phenomena related to the experiment. Since heat is deposited on the surface of the inspected sample it will travel in all directions once it is inside the sample. Due to the conductivity nature of the CFRP, heat would travel at different speeds at different directions in the sample and consequently heating up different parts of the sample earlier than other parts at the same depth. Since the target image has sharp edges for the defects and in the infrared images the edges of the defect are not as sharp, an IoU of 1 is virtually impossible.

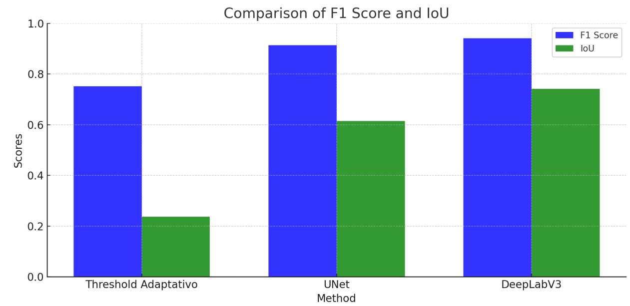


Figure 5. F1-Score and IoU.

The performance of the image segmentation techniques employed in this study was also compared with outcomes reported in similar studies. For instance, the research detailed in [21] utilized a U-Net architecture and achieved an F1-Score of 0.745 and an IoU of 0.593 for segmenting step-heating thermography images for composite laminates. This contrasts with our implementation of U-Net, which achieved a F1-Score of 0.9142 and an IoU of 0.6141, suggesting an enhanced segmentation capability possibly due the training methodologies and dataset variations. This same study applied the DeepLabv3 architecture in order to compare the results, they reported an F1-Score of 0.776 and an IoU of 0.629. Our results using DeepLabv3, with an F1-Score of 0.9406 and an IoU of 0.7419, surpass these figures, underscoring the effectiveness of our model's configuration and training process.

Authors in [22] reported a high IoU of 0.924 in their study on digital image correlation (DIC) based damage detection for CFRP laminates using machine learning-based image semantic segmentation, which is a distinct method compared to the infrared thermography employed in our work. While this study did not specify an F1-Score, the high IoU suggests extremely effective segmentation, which could be attributable to specialized processing techniques or highly optimized model parameters. While our IoU values are lower than those reported by [22], our method is tailored to handle the complexities of thermographic data, such as variable thermal contrasts and noise levels.

4. Conclusion

This study explored the application of image processing techniques and deep neural networks, specifically U-Net and DeepLabv3, for the segmentation of thermographic images of carbon fiber composites. The methods were evaluated based on metrics such as F1-Score and IoU, with DeepLabv3 demonstrating superior outcomes. This model achieved an F1-Score of 0.94 and an IoU of 0.74, highlighting its effectiveness in identifying structural defects in CFRP using infrared images from PT experiments.

The superiority of DeepLabv3 suggests that its approach is particularly suited for handling the complexities of thermographic images, providing a robust tool for NDT. This work not only underscores the efficiency of deep learning techniques in practical NDT applications but thus contributes with investigations on optimizing these models and their application in CFRP materials.

Although integrating infrared thermography with deep learning has shown promising results for defect segmentation in CFRP materials, certain challenges arise when applying this approach to real-world scenarios. The dataset used, while carefully curated and annotated, is relatively small and focuses on controlled scenarios with specific defect types and sizes. This constraint may limit the generalization of the models to more diverse or complex real-world conditions. Additionally, the experiments were conducted under controlled thermal imaging environments, which do not account for variations in ambient temperature, surface emissivity, or noise typically encountered in industrial settings. These factors could impact the accuracy of defect detection when applied in real-world scenarios. Moreover, the computational requirements of the deep learning models, particularly for real-time applications, present a challenge for scalability in industrial contexts.

Future work should address these limitations by incorporating larger, more diverse datasets, simulating or collecting data under varied conditions, and optimizing models for deployment in real-world systems. Future research could explore the generalization of the proposed approach to a range of composite materials beyond CFRP, such as Glass-Fibre-Reinforced Polymers (GFRP) or hybrid composites. This would involve testing the deep learning models on datasets that include varying material compositions, defect types, and thermal properties. Additionally, developing techniques to adapt the models for materials with different surface textures or emissivity characteristics could further enhance their applicability. Incorporating domain adaptation methods or transfer learning could also help address the challenges of limited labeled datasets for less-studied composite materials. Finally, real-world applications may benefit from integrating the approach with multimodal NDT techniques, such as ultrasonic or X-ray methods, to improve defect detection accuracy and reliability across diverse industrial scenarios.

Acknowledgements

This study was financed in part by the Coordenação de Aperfeicoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and by the National Council for Scientific and Technological Development - Brazil (CNPq) – Finance Codes 407140/2021-2 and 312530/2023-4. The authors thank the *Centro Multiusuário de Termografia Científica (CEMTEC)* for its support in the development of this work.

Author contributions

All authors have contributed equally to the development of this work.

References

[1] MALDAGUE, X. Theory and practice of infrared technology for nondestructive testing. 2001.

[2] MEOLA, C. Origin and theory of infrared thermography. In: ____. *Infrared Thermography Recent Advances and Future Trends*. [S.l.]: Bentham eBooks, 2012. p. 3–28.

[3] DONG, S.; WANG, P.; ABBAS, K. A survey on deep learning and its applications. *Computer Science Review*, v. 40, p. 100379, 2021.

[4] KYLILI, A. et al. Infrared thermography (irt) applications for building diagnostics: A review. *Applied Energy*, Elsevier, v. 134, p. 531–549, 2014.

[5] HEGDE, S.; SHENOY, B. S.; CHETHAN, K. Review on carbon fiber reinforced polymer (cfrp) and their mechanical performance. *Materials Today: Proceedings*, Elsevier, v. 19, p. 658–662, 2019.

[6] HEUER, H. et al. Review on quality assurance along the cfrp value chain–non-destructive testing of fabrics, preforms and cfrp by hf radio wave techniques. *Composites Part B: Engineering*, Elsevier, v. 77, p. 494–501, 2015.

[7] SWIDERSKI, W. Non-destructive testing of cfrp by laser excited thermography. *Composite Structures*, Elsevier, v. 209, p. 710–714, 2019.

[8] GUSTAFSSON, F.; GUSTAFSSON, F. *Adaptive filtering and change detection*. [S.l.]: Wiley New York, 2000. v. 1.

[9] RAJINIKANTH, V.; RAJA, N. S. M.; DEY, N. *A Beginner's Guide to Multilevel Image Thresholding*. [S.l.]: CRC Press, 2020.

[10] SPEAKMAN, J. R.; WARD, S. Infrared thermography: principles and applications. *Zoology-Jena-*, GUSTAV FISCHER, v. 101, p. 224–232, 1998.

[11] HE, Y. et al. Infrared machine vision and infrared thermography with deep learning: A review. *Infrared physics & technology*, Elsevier, v. 116, p. 103754, 2021.

[12] RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. In: SPRINGER. *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*. [S.l.], 2015. p. 234–241.

[13] DING, Y. et al. A stacked multi-connection simple reducing net for brain tumor segmentation. *IEEE Access*, IEEE, v. 7, p. 104011–104024, 2019.

[14] RAVIKUMAR, A.; SRIRAMAN, H. Acceleration of image processing and computer vision algorithms. In: *Handbook of Research on Computer Vision and Image*

Processing in the Deep Learning Era. [S.l.]: IGI Global, 2023. p. 1–18.

[15] CHEN, L.-C. et al. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017.

[16] PEDRAYES, O. D. et al. Semantic segmentation for non-destructive testing with step-heating thermography for composite laminates. *Measurement*, Elsevier, v. 200, p. 111653, 2022.

[17] SONAVANE, R. et al. Exploring resnet101, inceptionv3, and xception for modi script character classification. *International Journal of Intelligent Systems and Applications in Engineering*, v. 12, n. 17s, p. 117–124, 2024.

[18] FERNANDES, H. et al. Carbon fiber composites inspection and defect characterization using active infrared thermography: numerical simulations and experimental results. *Applied optics*, Optica Publishing Group, v. 55, n. 34, p. D46–D53, 2016.

[19] OBI, J. C. A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences*, World Journal of Advanced Engineering Technology and Sciences, v. 8, n. 1, p. 308–314, 2023.

[20] RAINIO, O.; TEUHO, J.; KLÉN, R. Evaluation metrics and statistical tests for machine learning. *Scientific Reports*, Nature Publishing Group UK London, v. 14, n. 1, p. 6086, 2024.

[21] PEDRAYES, O. D. et al. Semantic segmentation for non-destructive testing with step-heating thermography for composite laminates. *Measurement*, Elsevier, v. 200, p. 111653, 2022.

[22] WANG, Y. et al. Digital image correlation (dic) based damage detection for cfrp laminates by using machine learning based image semantic segmentation. *International Journal of Mechanical Sciences*, Elsevier, v. 230, p. 107529, 2022.