Towards Generalizable Crack Segmentation in Laser Thermography using Foundation Models

by B. Rahou[a], K. Helvig[a], P. Trouvé-Peloux[a], L. Gaverina[b], T. Sentenac[c]

[a] DTIS, ONERA - Université Paris-Saclay, Palaiseau, France
[b] DMAS, ONERA - Université Paris-Saclay, Châtillon, France
[c] Institut Clément Ader, Université de Toulouse, CNRS, IMT Mines Albi, INSA, ISAE-SUPAERO, UT3, Albi, France

Abstract

Flying-Spot Thermography (FST) detects surface cracks by monitoring thermal diffusion perturbations induced by cracks under a moving laser heat source. Despite the effectiveness of automated models in FST scenarios, adapting them to evolving pipelines with diverse materials and geometries requires continuous updates. However, sequential training triggers catastrophic forgetting in standard networks, overwriting past knowledge. To overcome this stability-plasticity dilemma, we propose a continual learning framework leveraging a DINOv2 vision foundation model. Evaluated across public and proprietary industrial datasets, our approach retains over 90% of baseline performance, decisively outperforming specialized CNNs.

1. Introduction

Among non-destructive testing (NDT) techniques, Flying-Spot Thermography (FST) has established itself as a reliable method for the inspection of surface and near-surface defects. The technique relies on a moving laser heat source to scan the surface of a target component. Local perturbations in heat diffusion around the laser spot, as revealed in the measurement of the surface temperature by an infrared camera, are an indication of the presence of defects. Initially introduced in the late 1960s for the inspection of aerospace components [1], FST offers a non-intrusive and efficient means of detecting defects without requiring complex instrumentation. While the physical acquisition of FST data is well-established, the automated analysis of these thermal sequences remains a significant challenge. The thermal signature of a crack typically appears as a subtle, high-frequency discontinuity in the temperature field, which can be difficult to distinguish from confounding factors such as geometric irregularities or measurement noise. In recent years, the state of the art in automated thermographic analysis has shifted toward deep learning, predominantly using Convolutional Neural Network (CNN) architectures trained to perform detection, localization, and even dense pixel-wise segmentation [2, 3].

While standard CNN-based methods (such as U-Net architectures [4]) achieve reliable performance in static industrial scenarios, they exhibit critical vulnerabilities when deployed in dynamic environments. Although many NDT pipelines operate under fixed parameters, real-world industrial lifecycles sometimes require inspection systems to adapt to domain shifts, such as novel composite structures or updated sensing hardware [5]. In complex tasks like crack segmentation, CNN models tend to learn local texture statistics and specific thermal contrasts rather than understanding the underlying structural semantics of the defects. Consequently, they generalize poorly to unseen experimental setups involving different camera sensors or material surfaces. Furthermore, if a standard CNN is fine-tuned on newly acquired data to accommodate these changes, it suffers from *catastrophic forgetting [6, 7]*: the network aggressively updates its weights, rapidly losing its ability to detect defects in the original domain. Since retraining models from scratch for every new setup is both computationally expensive and operationally impractical, static CNN-based approaches remain fundamentally limited for such evolving environments.

In the deep learning literature, adapting to such evolving environments is formalized as *continual learning* [8, 9]: a paradigm where a model is trained sequentially on new datasets without retaining access to previous ones. This requirement introduces the stability-plasticity dilemma: an inspection model must be plastic enough to learn new defect representations and thermal backgrounds, yet stable enough to preserve its previously acquired knowledge. To bridge this stability-plasticity gap, our present work investigates the use of Vision Foundation Models (FMs) for continual crack segmentation across multiple active thermography scenarios. Foundation models, such as DINOv2 [10], are deep neural networks pretrained on massive, highly diverse datasets to extract rich, generalizable representations. This massive exposure gives them a robust global understanding of shape, structure, and texture, allowing them to capture higher-level semantics rather than relying on local texture cues [11]. We propose an architecture composed of a pretrained DINOv2 encoder coupled with a Fully Convolutional Network (FCN) decoder. To validate this approach under a continual learning framework, we systematically compare our approach against standard convolutional and transformer models using two distinct laser thermography datasets of different domains (different structure and material). We evaluate joint-dataset performance, zero-shot generalization (performance on unseen datasets), transfer and resistance to forgetting. The main contributions of this study are as follows:

- **Application of Foundation Models to NDT:** Introducing the application of vision foundation models (DINOv2) for

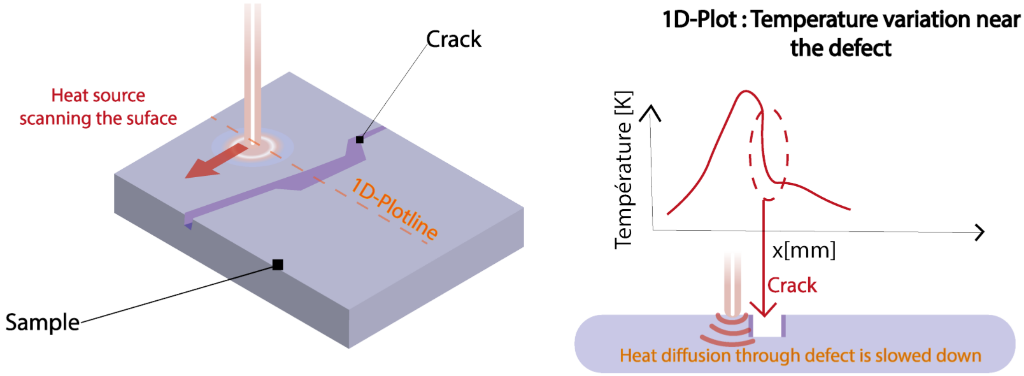


Fig. 1. *Schematic principle of Flying-Spot Thermography (FST).*

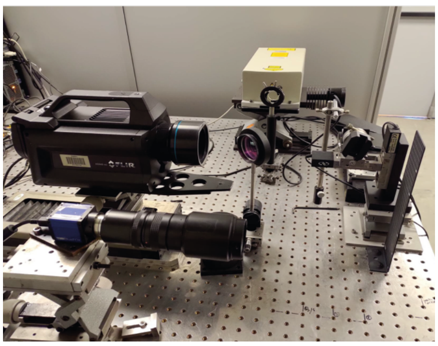
active thermography segmentation.

- **Continual Learning Benchmark:** Formulating a continual learning setup for laser thermography, systematically evaluating joint-dataset performance, zero-shot generalization, transfer and resistance to forgetting.
- **Evaluation of the Proposed Approach:** Demonstrating that the proposed DINOv2+FCN architecture retains over 90% of its original performance during sequential domain adaptation, whereas standard CNN baselines collapse to near-zero accuracy.
- **Dataset Contribution:** Introducing a novel publicly available thermography segmentation dataset.

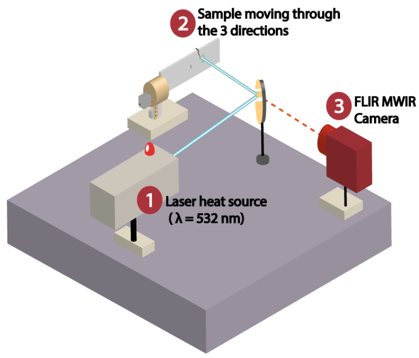
2. Experimental Setup and Dataset Contributions

2.1 Theoretical Principle and FST Bench

The principle of Flying-Spot Thermography (FST) relies on a localized moving laser heat source scanning the surface of a sample. A surface-breaking defect acts as a thermal barrier, producing a measurable discontinuity in the transient heat diffusion that is captured by an infrared camera. This process is illustrated in Figure 1. In the FST state of the art, the material is typically scanned back and forth to subtract the reconstructed thermogram from its return scan. This physical pre-processing cleans the data from structural noise and edge effects. However, to significantly reduce the examination duration, we use a single-pass scan. The processing of this data is subsequently more challenging, as it retains local gradients from structural artifacts that can easily be misclassified as cracks, thereby justifying the use of high-level deep learning architectures. We suppose the thermal emissivity and diffusivity to be constant, given the small temperature variations induced by the low-power heat source in the Mid-Wave Infrared (MW-IR) bandwidth.



(a) *Physical FST bench setup at ONERA.*



(b) *Optical path schematic.*

Fig. 2. *Experimental Flying-Spot Thermography setup.*

The experimental data for all datasets was acquired using the custom FST bench at ONERA, as illustrated in Figure 2. Heating is performed using a continuous laser source with a power varying between 0.5 and 3 W at a wavelength of 532 nm. A dichroic lens is employed to reflect the laser spot onto the part while transmitting the emitted infrared heat flux to the MW-IR camera, which is sensitive between 3 and 5 μm. The laser spot radius is 1.5 mm. During the scan, the spot covers a distance of 4.5 mm for horizontal scans and 6 mm for vertical scans in the vicinity of the defect, with translation velocities ranging from 0.5 to 1.5 mm/s.

2.2 Dataset Domains for Continual Learning

To evaluate the stability-plasticity dilemma in a realistic NDT lifecycle, we formulate our continual learning benchmark across two distinct physical domains. These datasets introduce significant variations in material properties, sample geometry, and thermal diffusion behavior. Representative examples illustrating the two domains are provided in Figure 3:

- **Domain A (Dataset Flyd):** This dataset consists of six rectangular pure aluminum test specimens presenting homogeneous surface properties. The material properties are well-characterized, with a thermal conductivity of $K = 237\ \mathrm{W/(m \cdot K)}$ and a thermal diffusivity of $\alpha = 9.7 \times 10^{-5}\ \mathrm{m^2/s}$. While the raw thermographic sequences were originally introduced by Helvig et al. [12] for defect localization, we have manually added pixel-level ground truth masks to adapt this repository for semantic crack segmentation tasks. This newly structured dataset is introduced as a public contribution to the NDT community and is openly available at https://github.com/kevinhelvig/FLYD.
- **Domain B (Superalloy):** To simulate a complex industrial inspection scenario, this dataset comprises superalloy samples coated with a thermal barrier. The introduction of the thermal barrier coating fundamentally alters the surface emissivity and heat diffusion compared to pure aluminum, presenting a highly realistic domain shift. For confidentiality reasons, this dataset cannot be made public.

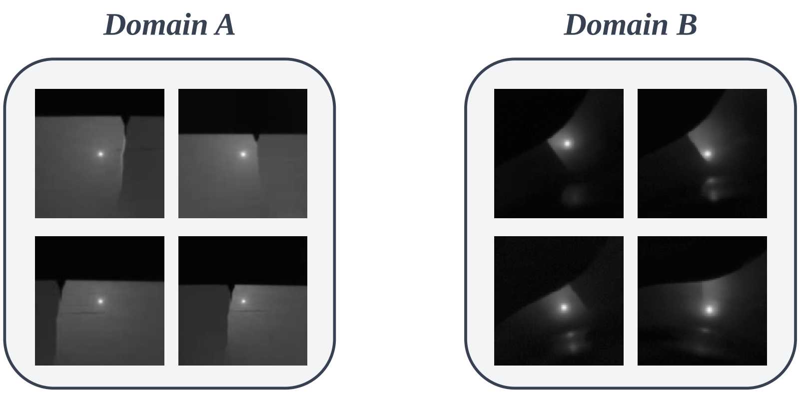


Fig. 3. *Visual comparison of the evaluated datasets. (Left) Domain A: pure aluminum specimens from the Flyd dataset presenting homogeneous surface characteristics. (Right) Domain B: coated superalloy samples.*

3. Methodology: Foundation Models for Continual Segmentation

3.1 Continual Learning Framework

In an industrial Non-Destructive Testing (NDT) environment, inspection parameters such as materials, surface geometries, and sensing hardware may evolve. To bridge the stability-plasticity gap inherent in such dynamic scenarios, this work investigates the use of Vision Foundation Models (FMs) for continual crack segmentation across multiple active thermography settings. We formulate a continual learning framework using the two distinct laser thermography datasets described in Section 2. The evaluation protocol is structured around four key paradigms:

- **Joint-Dataset Performance:** The model is trained and evaluated simultaneously on all available domains. This establishes a multi-task upper-bound baseline, which reveals the maximum performance a model can theoretically achieve when granted unrestricted access to all domains without the constraint of sequential data presentation.

- **Zero-Shot Generalization:** The model is evaluated on a entirely unseen dataset without any prior training, testing the inherent universality of the extracted thermal features.

- **Transfer Learning:** The model is fine-tuned on a new domain to evaluate its plasticity and efficiency when adapting to novel NDT inspection conditions.
- **Resistance to Forgetting:** After a transfer (e.g., Domain A then Domain B), we evaluate the performance on the initial domain to quantify retention abilities.

3.2 Proposed Architecture: DINOv2 and FCN Decoder

To effectively adapt to challenging domain shifts, we propose a straightforward yet highly effective segmentation architecture. The network is composed of a pretrained DINOv2 encoder directly coupled with a Fully Convolutional Network (FCN) decoder, as illustrated Figure 4.

Unlike conventional Convolutional Neural Networks (CNNs) that rely predominantly on local texture cues and specific thermal contrasts, foundation models such as DINOv2 are deep neural networks pretrained on massive, highly diverse datasets. This extensive exposure gives to the DINOv2 backbone a robust global understanding of shape, structural topology, and thermal gradients. By leveraging these rich, generalizable representations, the encoder captures the higher-level semantics of a crack, effectively distinguishing actual defect signatures from benign surface variations or geometric edge effects.

The high-level spatial features extracted by the DINOv2 backbone are fed into the lightweight FCN decoder. The decoder's role is to progressively upsample these semantic representations back to the original image resolution, generating a precise, pixel-wise segmentation mask of the defect. By maintaining architectural simplicity, we isolate and demonstrate the inherent representational power of the foundation model, providing an accessible and highly adaptable framework for automated continual thermographic analysis.

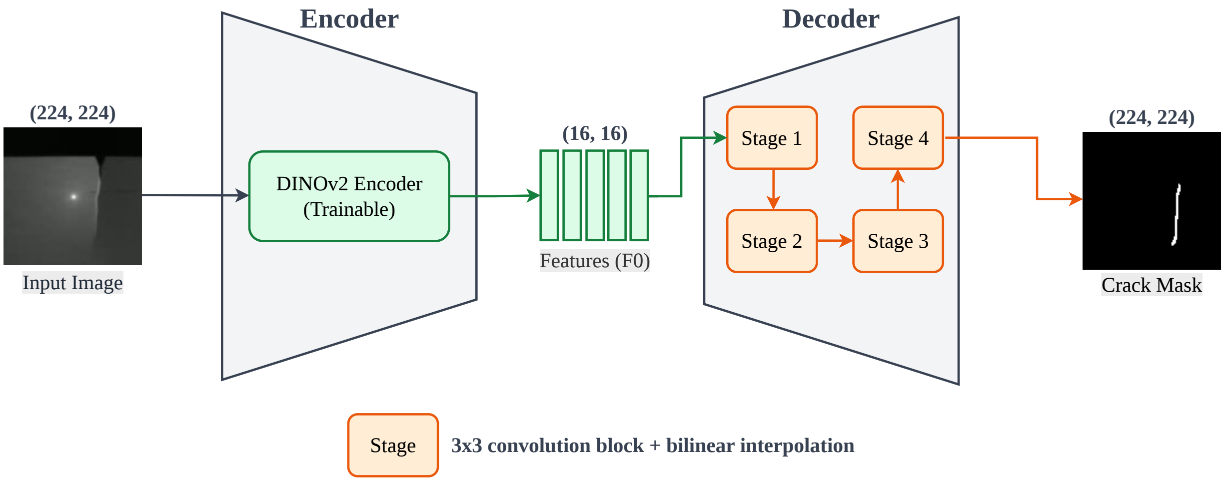


Fig. 4. *Schematic of the proposed DINOv2+FCN architecture. The input thermal image is processed by the pretrained DINOv2 encoder, which extracts high-dimensional, generalizable semantic features (16 x 16 tokens). These representations are then processed by the FCN decoder, progressively upsampling the latent space to generate a full-resolution segmentation mask.*

4. Experiments

4.1 Evaluation Protocol

To rigorously assess the models across the proposed continual learning framework, the datasets were carefully split to prevent data leakage between the training and testing phases. For Domain A (the Flyd dataset, featuring pure aluminum), the collection consists of six physical samples. Four of these samples were designated for the training phase (subsequently partitioned into an 80% training and 20% validation split), while the remaining two samples were kept entirely isolated to serve as the testing set.

Domain B (the coated superalloy), coming from a highly distinct industrial environment, comprises thirteen physical samples: six containing localized cracks and seven defect-free control samples. To ensure balanced representation, four cracked and five defect-free samples were allocated to construct the training corpus, which was also subjected to an

80%/20% train-validation split. The remaining four samples (two cracked and two defect-free) were strictly reserved for the testing set. A summary of the dataset distributions is provided in Table 1.

Table 1. *Dataset Distribution and Split Overview.*

Domain	Material	Total Samples	Total images
Domain A (Flyd)	Pure Aluminum	6 (4 train, 2 test)	701
Domain B	Coated Superalloy	13 (9 train, 4 test)	696

To provide a comprehensive benchmark against the proposed DINOv2+FCN approach, three established architectures were selected as baselines, representing the evolution of deep learning for crack segmentation. **U-Net [4]** serves as the foundational standard for dense pixel-wise segmentation. **DeepCrack [13]** is included as a specialized convolutional architecture explicitly designed for crack detection. Finally, **CrackFormer [14]** represents the current state-of-the-art in specialized non-convolutional crack segmentation, leveraging a transformer architecture.

The baseline models (UNet, Deepcrack and Crackformer) were optimized following the standard hyperparameter configurations, learning rates, and batch sizes specified in their respective original publications to ensure a fair and representative benchmark. Conversely, our proposed DINOv2+FCN architecture was optimized via the AdamW optimizer [15]. The pretrained backbone (`facebook/dinov2-base`) was finetuned with a highly conservative learning rate of $\eta_{\text{encoder}} = 1 \times 10^{-6}$ to preserve its robust foundation representations and shield them from catastrophic forgetting. Concurrently, the newly initialized FCN decoder head was allocated a standard learning rate of $\eta_{\text{decoder}} = 1 \times 10^{-3}$ to allow rapid, expressive alignment to the target crack segmentation masks. All architectures in this study were trained using an Nvidia A5000 GPU setup for a maximum of 100 epochs, with an early stopping criterion implemented to halt training if the validation loss did not improve for 20 consecutive epochs.

In practice, the evaluation protocol was executed systematically across three stages for each of the four architectures: first, the model was trained on the pure aluminum dataset (Domain A) to establish the *In-Domain* baseline and immediate *Zero-Shot* generalization capabilities on the unseen coated superalloy data. Second, this same model was sequentially fine-tuned on Domain B to measure its *Transfer* performance, followed by an immediate backward evaluation on Domain A to conduct the *Retention* check. Finally, a separate iteration of each model was trained simultaneously on both datasets to evaluate the multi-domain *Joint* upper bound.

To quantify pixel-wise crack segmentation accuracy across these phases, we report four standard metrics. The mean Intersection over Union (*Mean IoU*) is computed to track overall pixel-wise structural alignment. Classification accuracy is evaluated using *Precision*, which is the proportion of true positives among all predicted crack pixels, and *Recall*, which quantifies the fraction of the physical ground truth successfully detected. Finally, the *F1 Score* serves as our primary comparative metric: as the harmonic mean of Precision and Recall, it ensures a balanced performance assessment despite the extreme class imbalance native to sparse defect masks.

4.2 Continual Learning Results and Discussion

The quantitative metrics acquired from our sequential domain adaptation benchmark, transitioning from the pure aluminum baseline (Domain A) to the coated superalloy dataset (Domain B), are aggregated comprehensively across all evaluated architectures in Table 2.

Table 2. *Quantitative performance comparison across the sequential framework ($A \rightarrow B$) in percentage (%).*

Model	Metric	In-Domain (A)	Zero-Shot (B)	Transfer (B)	Retention (A)	Joint (A / B)
CrackFormer	IoU (Mean)	65.45	49.87	67.01	53.29	66.01 / 60.60
	Precision	40.18	0.49	42.14	19.90	42.49 / 41.36
	Recall	58.40	**10.97**	63.94	9.61	57.53 / 30.38
	F1 Score	47.61	0.95	50.80	12.96	48.88 / 35.03
U-Net	IoU (Mean)	68.60	**49.98**	67.29	52.30	68.17 / 72.50
	Precision	49.28	0.00	55.64	13.49	48.00 / **63.14**
	Recall	60.98	0.00	47.78	7.29	60.70 / 61.09
	F1 Score	54.51	0.00	51.41	0.95	53.61 / 62.10
DeepCrack	IoU (Mean)	67.78	**49.98**	63.13	62.59	67.09 / 60.04
	Precision	45.93	0.00	39.72	36.17	44.25 / 35.65
	Recall	62.04	0.00	43.77	46.52	60.99 / 31.59
	F1 Score	52.78	0.00	41.65	40.70	51.29 / 33.50
DINOv2 + FCN (Ours)	**IoU (Mean)**	**70.53**	**49.98**	**72.93**	**68.60**	**73.23** / **73.72**
	Precision	**52.70**	0.00	**60.80**	**63.23**	**57.39** / 60.15
	Recall	**65.64**	0.00	**65.16**	**47.83**	**71.45** / 69.23
	F1 Score	**58.47**	0.00	**62.90**	**54.46**	**63.65** / **64.37**

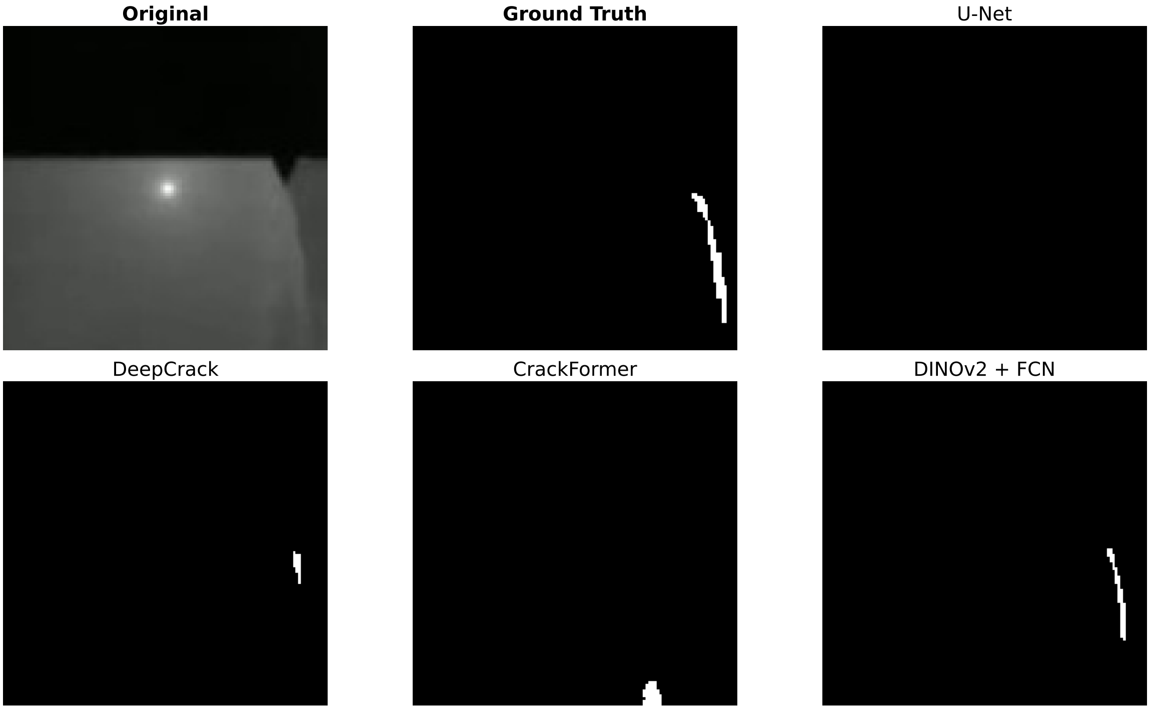


Fig. 5. *Qualitative comparison of Retention performance. The figure displays the ground truth alongside predictions from the baselines and from our approach evaluated on Domain A after learning domain B.*

Analysis of Isolated Baselines and the Zero-Shot Domain Collapse Initial training exclusively on Domain A (*In-Domain*) reveals that all deep models possess the capability to extract crack geometry from pure aluminum. Under static conditions, our proposed DINOv2+FCN architecture yields the strongest starting performance, reaching an F_1 score of 58.47%, followed by U-Net (54.51%) and DeepCrack (52.78%).

However, evaluating these Domain-A-optimized models directly on Domain B without any transfer or learning on B (*Zero-Shot* phase) highlights a severe performance collapse across all architectures. The U-Net and DeepCrack models fail completely, dropping to an absolute F_1 score of 0.00%, while CrackFormer outputs a negligible and noisy F_1 score of 0.95%. Our proposed DINOv2+FCN architecture similarly bottoms out at 0.00%. This total blindness to cracks is reflected by a systematic *IoU Mean* $\approx 50\%$ across all models. This occurs because the crack IoU drops to absolute zero while the background IoU approaches 100% due to the model always predicting background for every pixel of the dataset. From a computer vision perspective, this universal collapse exposes the severity of the underlying domain shift between the two datasets. From a thermographic point of view, this is explained by the fundamental differences in materials between the aluminum and the superalloy. Specifically, the superalloy in Domain B features a thermal barrier coating that makes the thermographic inspection significantly more difficult and results in a much worse thermal contrast compared to the bare aluminum.

Transfer Efficiency and Catastrophic Forgetting When the models undergo sequential fine-tuning on the coated superalloy dataset (*Transfer* phase), distinct structural behaviors emerge. The DINOv2+FCN architecture demonstrates excellent plasticity, rapidly converging to a top-tier F_1 score of 62.90%. This outpaces U-Net (51.41%) and CrackFormer (50.80%), whereas DeepCrack exhibits restricted adaptation capacity, stalling at an F_1 score of 41.65%.

The true operational difference, however, is revealed during the *Retention* phase, where the newly adapted models are evaluated back on Domain A without any further weight adjustments. Here, the baselines suffer an explicit structural failure. The U-Net's detection capability collapses entirely, dropping from its 54.51% baseline down to a near-zero F_1 score of 0.95%. CrackFormer experiences a parallel collapse, retaining an F_1 score of only 12.96%. This collapse is visually mapped in Figure 5, which highlights how a transferred U-Net aggressively scrubs away its previous weights to accommodate Domain B's textures, leaving it unable to distinguish cracks back on Domain A. While DeepCrack appears to resist this drop (F_1: 40.70%), it does so by compromising its plasticity: its weights remain locked in the original domain, resulting in its poor performance on the new superalloy targets (41.65%).

In contrast, our DINOv2+FCN model handles the stability-plasticity dilemma successfully. When evaluated back on Domain A, it preserves an F_1 score of 54.46%. This indicates that our model maintains 93.1% of its original baseline segmentation performance ($54.46/58.47$).

Cross-Domain Insights from Joint Training Within standard continual learning protocols, simultaneous joint training is used to establish a multi-task upper bound. By entirely removing the constraint of sequential data presentation, it reveals the maximum performance a model can theoretically achieve when granted unrestricted access to all domains without the risk of catastrophic forgetting. However, the final column of Table 2 reveals that the baseline architectures fail to capitalize on this ideal scenario. For specialized models like CrackFormer and DeepCrack, simultaneous exposure to both datasets actually harms performance, dropping their scores below their isolated single-domain baselines. This drop is driven by the massive domain gap between the datasets.

In contrast, our DINOv2+FCN approach demonstrates a powerful cross-domain synergy. When trained jointly, its performance rises to an F_1 score of 63.65% on Domain A and 64.37% on Domain B, outperforming its isolated single-domain baselines. We hypothesize that because DINOv2 leverages robust, large-scale pre-trained representations, it remains insulated from task interference when exposed to complex, multi-material data. Instead, it uses the simultaneous exposure to learn a broader, unified physical representation of thermal boundaries. These results show that while specialized models, whether convolutional or transformer-based, can successfully learn static FST inspections, pre-trained vision foundation models provide the robust, generalizable framework required for automated industrial thermography lines that must seamlessly adapt across materials.

5. Conclusion

In this work, we addressed the challenge of domain shift and catastrophic forgetting in automated Flying-Spot Thermography (FST) inspections. While deep learning has proven highly capable of extracting complex defect geometries from localized thermal gradients, traditional architectures fail when operational conditions change. To address this, we developed a dedicated FST defect segmentation dataset on pure aluminum, providing a rigorous benchmark for evaluating model stability. As demonstrated by our sequential adaptation benchmark ($A \to B$), transitioning from this highly conductive surface to the restrictive thermal barrier of a coated superalloy triggers a universal zero-shot collapse. Additionally, when forced to sequentially adapt to this new material domain, standard convolutional and transformer-based baselines, including U-Net, DeepCrack, and CrackFormer, suffered from catastrophic forgetting, effectively overwriting their structural knowledge of the original domain.

To overcome this stability-plasticity dilemma, we proposed a novel continual learning framework leveraging a DINOv2 vision foundation backbone paired with an adaptable Fully Convolutional Network (FCN) decoder. Our approach successfully retained 93.1% of its original baseline performance on pure aluminum after fully adapting to the coated superalloy. Furthermore, our evaluation of the multi-task joint upper bound revealed that while specialized models suffer from negative task interference when exposed to conflicting material physics, the DINOv2 foundation model achieves positive cross-domain transfer, using the diverse exposure to build a more generalized understanding of thermal defect boundaries.

Building upon these findings, future work will aim to extend this continual learning protocol across a broader sequence of distinct material domains (e.g., $A \to B \to C \to D$). Evaluating the framework on other diversified samples such as ceramics or composites will be crucial for analyzing the long-term memory capacity of the vision foundation model and assessing its limits for heterogeneous real-world deployments.

References

[1] Edward J. Kubiak. Infrared detection of fatigue cracks and other near-surface defects. *Applied Optics*, 7(9):1743–1747, 1968.

[2] Kevin Helvig, Pauline Trouvé-Peloux, Ludovic Gaverina, Baptiste Abeloos, and Jean-Michel Roche. Automated crack detection on metallic materials with flying-spot thermography using deep learning and progressive training. *Quantitative InfraRed Thermography Journal*, 22:1–20, 2023.

[3] Qiang Fang, Clemente Ibarra-Castanedo, Iván Garrido, Yuxia Duan, and Xavier Maldague. Automatic detection and identification of defects by deep learning algorithms from pulsed thermography data. *Sensors*, 23(9):4444, 2023.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241, 2015.

[5] Yuqi Weng, Yahui Li, Xinyu Li, and Yanning Zhang. Unsupervised domain adaptation for crack detection. *Automation in Construction*, 153:104939, 2023.

[6] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan

Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

[7] Robert M. French. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3(4):128–135, 1999.

[8] German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019.

[9] Gido M. van de Ven and Andreas S. Tolias. Three scenarios for continual learning. *arXiv preprint arXiv:1904.07734*, 2019.

[10] Maxime Oquab and et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

[11] Muzammal Naseer, Kanchana Ranasinghe, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Intriguing properties of vision transformers. In *NIPS*, pages 23296–23308, 2021.

[12] K. Helvig, P. Trouvé-Peloux, L. Gaverina, J.-M. Roche, B. Abeloos, and C. Pradère. Laser flying-spot thermography: an open-access dataset for machine learning and deep learning. In *Proceedings of the Sixteenth International Conference on Quality Control by Artificial Vision*, pages 334–340. SPIE, 2023.

[13] Yahui Liu, Jian Yao, Xiaohu Lu, Renping Xie, and Li Li. Deepcrack: A deep hierarchical feature learning architecture for crack segmentation. *Neurocomputing*, 338:139–153, 2019.

[14] Hu Liu, Xinyu Miao, and Ajmal Mian. Crackformer-ii: A transformer-based network for fine-grained crack detection. *arXiv preprint arXiv:2308.15174*, 2023.

[15] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations*, 2019.