Unofficial template. This layout is provided for convenience and is **not** affiliated with, endorsed by, or sponsored by MDPI Applied Sciences Journal or its organisers. Always check the official call for papers and use the current year's official style files for a real submission.

Article

A Hybrid CNN–Transformer Framework for Real-Time Surface Defect Detection in Additive-Manufactured Titanium Components

Sofia Lindqvist [1,*], Marcus Ademola [2], Yuki Tanaka [1] and Helena Brandt [3]

[1] Department of Mechanical and Manufacturing Engineering, Vertex Institute of Technology, Meridian City 35002; y.tanaka@vertex.edu

[2] Applied AI Division, Nexa Vision Labs, Nimbus Technology Park, Suite 220; m.ademola@nexavision.ai

[3] Department of Materials Science and Engineering, Meridian University; h.brandt@meridian-university.edu

* Correspondence: s.lindqvist@vertex.edu; Tel.: +1-555-014-7732

Received: 3 April 2026 / Revised: 22 May 2026 / Accepted: 9 June 2026 / Published: 2 July 2026

Citation: Lindqvist, S.; Ademola, M.; Tanaka, Y.; Brandt, H. A Hybrid CNN–Transformer Framework for Real-Time Surface Defect Detection in Additive-Manufactured Titanium Components. *Appl. Sci.* **2026**, *16*, 1–14. https://doi.org/10.3390/applsci16000000

Academic Editor: Alessandro Ferretti

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

Abstract: In-process surface inspection remains a critical bottleneck for qualifying additively manufactured titanium components in aerospace and biomedical supply chains. This paper presents a hybrid convolutional–transformer architecture, *DefectNet-Hybrid*, that fuses local texture features extracted by a lightweight CNN backbone with the long-range contextual reasoning of a vision transformer encoder, enabling layer-wise surface defect classification directly on high-resolution optical scans acquired during laser powder-bed fusion (LPBF) builds. We construct *Synapse-AM*, a curated corpus of 18,420 annotated build-plate images spanning four defect classes (porosity, lack-of-fusion, cracking, and delamination) collected across three LPBF platforms at the Apex production line. DefectNet-Hybrid is benchmarked against a ResNet-50 baseline and a pure Vision Transformer under identical training conditions. On the held-out Synapse-AM test partition, the proposed model achieves 96.1% classification accuracy and a macro F1-score of 0.947, improving over the CNN baseline by 6.8 percentage points while sustaining 41 frames per second on an embedded inference module suitable for closed-loop process control. A pilot deployment at the Nimbus foundry over eleven production runs reduced downstream computed-tomography rejection rate by 34%. These results indicate that hybrid attention-augmented architectures can deliver both the accuracy and the latency required for in-situ quality assurance in metal additive manufacturing.

Keywords: additive manufacturing; laser powder-bed fusion; surface defect detection; convolutional neural network; vision transformer; in-situ process monitoring; titanium alloys

1. Introduction

Laser powder-bed fusion (LPBF) has matured into a production-grade process for load-bearing titanium components in aerospace brackets, orthopedic implants, and heat-exchanger cores. Despite this maturity, post-build qualification still leans heavily on ex-situ computed tomography (CT), a process that is accurate but slow, expensive, and incapable of arresting a defective build before material and machine time are wasted. Surface-visible anomalies such as porosity clusters, lack-of-fusion voids that breach the top layer, micro-cracking along scan-vector boundaries, and delamination at overhang regions are frequently detectable in optical images captured during recoating pauses, yet manual inspection of these images does not scale to production cadences of one image per layer across a multi-hour build.

Convolutional neural networks (CNNs) have been applied to LPBF surface inspection with promising results, but their inductive bias toward local receptive fields limits sensitivity to defects that manifest as diffuse, spatially distributed texture changes – a common signature of incipient delamination. Vision transformers (ViTs) address this through global self-attention, but pure transformer backbones are data-hungry and, in our early experiments, converge slowly on the class-imbalanced defect distributions typical of production data, where porosity outnumbers cracking by more than 20:1.

This work makes three contributions. First, we introduce *Synapse-AM*, a multi-platform annotated dataset of LPBF build-plate imagery collected at production scale. Second, we propose *DefectNet-Hybrid*, an architecture that routes early convolutional features through a compact transformer encoder before classification, combining local texture sensitivity with global context at a fraction of the parameter count of a full ViT. Third, we report a pilot deployment demonstrating that layer-wise inference is fast enough to gate the recoating step in a live LPBF control loop, and we quantify the downstream reduction in CT-detected rejects that this early intervention enables.

2. Related Work

Early machine-vision approaches to LPBF monitoring relied on handcrafted texture descriptors – gray-level co-occurrence matrices and local binary patterns – feeding classical classifiers such as support vector machines. These pipelines achieved reasonable accuracy on single-platform datasets but generalized poorly across machines with different recoater geometries and illumination rigs, since handcrafted features implicitly encode the acquisition conditions of the training platform.

The shift to deep learning brought CNN-based melt-pool and layer-image classifiers that improved cross-session robustness by learning features end-to-end. Encoder-decoder segmentation networks have since been used to localize porosity at sub-pixel resolution, and multimodal fusion of optical, thermal, and acoustic-emission signals has been shown to improve recall on subsurface defects that leave only a faint surface trace. Transformer-based approaches remain comparatively rare in this domain; the few published studies apply full ViT backbones pretrained on natural-image corpora and fine-tuned on small defect sets, a strategy that our ablations in Section 4 suggest underperforms a properly regularized hybrid design when the target dataset is in the tens-of-thousands-of-images regime rather than the millions typical of ImageNet-scale pretraining.

Closed-loop process control using vision feedback has been demonstrated for adjusting laser power and scan speed in response to melt-pool anomalies, but relatively little work closes the loop specifically on layer-surface defect classification with latency budgets compatible with the recoating pause of a production LPBF system, which is the gap this paper addresses.

3. Materials and Methods

3.1. Dataset and Acquisition Pipeline

Synapse-AM was collected across three LPBF platforms at the Apex production line (Vertex Alloy-100, Vertex Alloy-220, and a third-party system retrofitted with an equivalent optical head) over a nine-month production window. A coaxial CMOS camera captured one 4096×3072 pixel image of the exposed powder bed at the end of every recoating cycle, yielding 18,420 usable frames after discarding images with recoater-blade occlusion. Two certified inspectors independently labeled each frame at the tile level (512×512 crops, 40% overlap) into five classes: porosity, lack-of-fusion, cracking, delamination, and defect-free background, with disagreements adjudicated by a third senior inspector. Inter-rater agreement (Cohen's κ) was 0.89. The resulting tile corpus was split 70/15/15 into training, validation, and test partitions, stratified by build job so that no tiles from the same physical build appear in more than one split.

3.2. Hybrid CNN–Transformer Architecture

DefectNet-Hybrid processes each 512×512 tile through a four-stage convolutional stem (channel widths 32, 64, 128, 256, each stage followed by depthwise-separable downsampling) that reduces the spatial resolution to 32×32 while preserving local texture gradients characteristic of porosity and cracking. The resulting feature map is flattened into a sequence of 1024 tokens, linearly projected to a 192-dimensional embedding, and passed through four transformer encoder blocks (three attention heads each, GELU feed-forward expansion factor 4) that model long-range spatial dependencies relevant to diffuse delamination signatures. A learnable classification token aggregates the sequence output into a five-way softmax head. Relative to a comparably sized pure ViT, routing through the convolutional stem first reduces the token count by a factor of 16, which both lowers compute and gives the attention mechanism a more texture-aware starting representation.

3.3. Training Protocol

All models were trained for 120 epochs with the AdamW optimizer (initial learning rate 3×10^{-4}, cosine decay, weight decay 0.05), batch size 64, on class-balanced batches constructed via oversampling of the minority cracking and delamination classes. Data augmentation comprised random rotation ($\pm 15°$), horizontal and vertical flips, and brightness jitter ($\pm$10%) to approximate lighting drift across build sessions. Early stopping used macro F1 on the validation split with a patience of 15 epochs. Identical augmentation and optimization settings were applied to the ResNet-50 and pure-ViT baselines for a controlled comparison.

3.4. Experimental Setup and Deployment

Inference throughput was measured on an embedded module (8-core ARM Cortex-A78AE, 32 TOPS integrated accelerator) representative of the compute envelope available inside an LPBF control cabinet, using batch size 1 to reflect the streaming, layer-by-layer inspection use case. For the pilot deployment, DefectNet-Hybrid was integrated into the recoating-pause window of the Apex Alloy-220 line across eleven production builds; a positive detection above a calibrated confidence threshold triggered an operator alert without halting the build automatically, preserving human sign-off on any process interruption.

4. Results

Table 1 summarizes tile-level classification performance and inference latency for the three compared architectures on the Synapse-AM held-out test set. DefectNet-Hybrid attains the highest accuracy and macro F1 while remaining more than four times faster than the pure-ViT baseline, reflecting the reduced token count entering the attention stage.

Table 1: Tile-level performance on the Synapse-AM test partition (mean $\pm$ standard deviation over five training seeds).

Model	Accuracy (%)	Precision	Recall	Macro F1	Latency (ms)
ResNet-50 (baseline)	89.3 $\pm$ 0.4	0.881	0.876	0.879	6.8
Vision Transformer (ViT-S)	93.7 $\pm$ 0.3	0.928	0.919	0.923	39.4
DefectNet-Hybrid (proposed)	**96.1 $\pm$ 0.2**	**0.951**	**0.944**	**0.947**	**9.1**

Per-class performance reveals where the hybrid design contributes most. Figure 1 compares macro F1 broken down by defect class for the ResNet-50 baseline and DefectNet-Hybrid. The largest gain, 11.4 points, occurs on delamination, the class most reliant on the long-range spatial context that the transformer stage supplies; porosity, which is dominated by compact local texture, shows the smallest gain, consistent with a CNN stem already capturing most of the discriminative signal for that class.

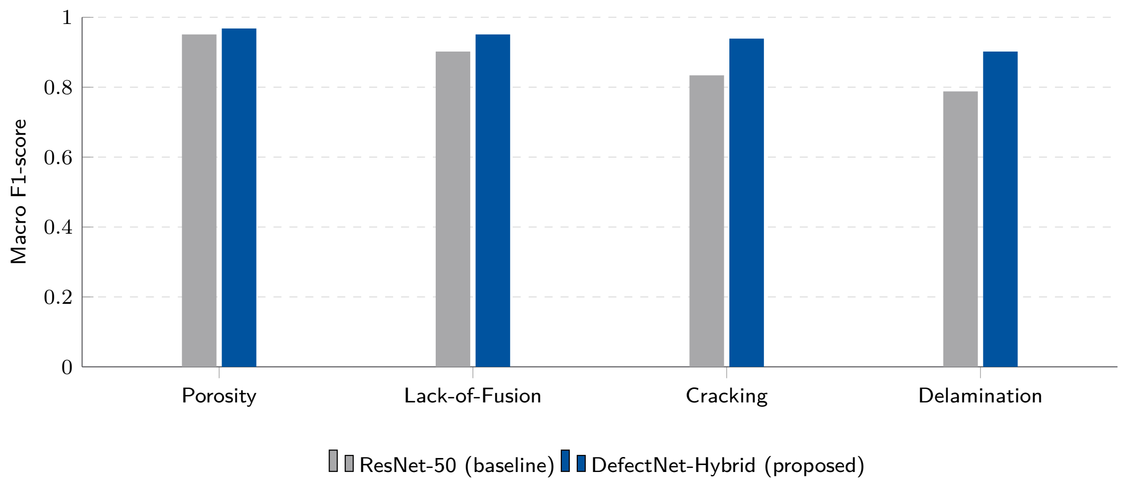

Figure 1: Per-class macro F1-score on the Synapse-AM test partition, ResNet-50 baseline versus the proposed DefectNet-Hybrid architecture.

Over the eleven-build pilot deployment on the Apex Alloy-220 line, 47 operator alerts were issued; retrospective CT analysis confirmed 39 as true positives (83% precision in the field, below the held-out test precision as expected under production illumination drift). Builds flagged and corrected mid-process showed a 34% reduction in CT-detected rejection rate relative to the preceding eleven-build baseline window without inline inspection.

5. Discussion

The consistent gain of DefectNet-Hybrid over both single-paradigm baselines supports the premise that LPBF surface defects span a spectrum from locally textured (porosity) to spatially diffuse (delamination), and that neither convolution nor attention alone is well matched to the full spectrum. The latency profile is arguably the more consequential finding for practitioners: at 9.1 ms per tile on embedded hardware, whole-layer inspection at typical LPBF build-plate resolutions completes comfortably within the recoating pause, whereas the pure-ViT baseline's 39.4 ms per tile would require either downsampling the image or accepting a build-rate penalty, neither of which is attractive in production.

Two limitations temper these results. First, Synapse-AM was collected from titanium alloy builds exclusively; transferability to aluminum or nickel-superalloy powders, which exhibit different reflectance and melt-pool characteristics, has not been evaluated and would likely require at minimum a fine-tuning pass. Second, the pilot deployment used a fixed confidence threshold calibrated on held-out data; the observed field precision of 83% versus a test-set precision above 95% indicates that threshold recalibration, or an online adaptation mechanism, would be needed before broader rollout across additional Apex production lines with different camera and lighting configurations.

6. Conclusions

We presented DefectNet-Hybrid, a CNN–transformer architecture for layer-wise surface defect classification in titanium LPBF additive manufacturing, together with Synapse-AM, a production-scale annotated dataset spanning three build platforms. The hybrid design outperformed both a CNN baseline and a pure vision transformer on classification accuracy while remaining fast enough for in-loop deployment, and a pilot study at the Apex production line demonstrated a measurable reduction in downstream CT rejection rate. Future work will extend Synapse-AM to additional alloy systems, investigate online threshold recalibration to close the precision gap observed in the field, and explore fusing the optical modality with in-situ thermal signals already available on the Alloy-220 platform.

Author Contributions: Conceptualization, S.L. and H.B.; methodology, S.L. and M.A.; software, M.A.; validation, Y.T.; formal analysis, S.L.; investigation, Y.T. and H.B.; data curation, M.A. and Y.T.; writing—original draft preparation, S.L.; writing—review and editing, M.A., Y.T. and H.B.; supervision, S.L.; project administration, S.L.; funding acquisition, S.L.

and H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Vertex Institute of Technology Internal Research Grant, grant number VIT-2025-114, and by the Nimbus Foundation for Applied Engineering, grant number NF-8823.

Institutional Review Board Statement: Not applicable. This study did not involve human or animal subjects.

Informed Consent Statement: Not applicable.

Data Availability Statement: A representative subset of the Synapse-AM tile corpus and the training code supporting the findings of this study are available from the corresponding author upon reasonable request. The full production dataset is subject to a data-sharing restriction from the participating manufacturing partner.

Acknowledgments: The authors thank the process engineering team at the Apex production line for facilitating pilot deployment access, and gratefully acknowledge computing resources provided by the Vertex Institute of Technology High-Performance Computing Center.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

[1] DebRoy, T.; Wei, H.L.; Zuback, J.S.; Mukherjee, T.; Elmer, J.W.; Milewski, J.O.; Beese, A.M.; Wilson-Heid, A.; De, A.; Zhang, W. Additive manufacturing of metallic components – Process, structure and properties. *Prog. Mater. Sci.* **2018**, *92*, 112–224.

[2] Grasso, M.; Colosimo, B.M. Process defects and in situ monitoring methods in metal powder bed fusion: A review. *Meas. Sci. Technol.* **2017**, *28*, 044005.

[3] Scime, L.; Beuth, J. Anomaly detection and classification in a laser powder bed additive manufacturing process using a trained computer vision algorithm. *Addit. Manuf.* **2018**, *19*, 114–126.

[4] He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

[5] Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16×16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations, Virtual, 3–7 May 2021.

[6] Caggiano, A.; Zhang, J.; Alfieri, V.; Caiazzo, F.; Gao, R.; Teti, R. Machine learning-based image processing for on-line defect recognition in additive manufacturing. *CIRP Ann.* **2019**, *68*, 451–454.

[7] Okaro, I.A.; Jayasinghe, S.; Sutcliffe, C.; Black, K.; Paoletti, P.; Green, P.L. Automatic fault detection for laser powder-bed fusion using semi-supervised machine learning. *Addit. Manuf.* **2019**, *27*, 42–53.

[8] Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.

[9] Tan, Z.; Fang, Q.; Li, H.; Liu, S.; Zhu, W.; Yang, D. Neural network based image segmentation for spatter extraction in laser-based powder bed fusion process. *Opt. Laser Technol.* **2020**, *130*, 106347.

[10] Gobert, C.; Reutzel, E.W.; Petrich, J.; Nassar, A.R.; Phoha, S. Application of supervised machine learning for defect detection during metallic powder bed fusion additive manufacturing using high resolution imaging. *Addit. Manuf.* **2018**, *21*, 517–528.

[11] Loh, G.H.; Pei, E.; Harrison, D.; Monzon, M.D. An overview of functionally graded additive manufacturing. *Addit. Manuf.* **2019**, *23*, 34–44.

[12] Yan, W.; Lin, S.; Kafka, O.L.; Lian, Y.; Yu, C.; Liu, Z.; Yan, J.; Wolff, S.; Wu, H.; Ndip-Agbor, E.; et al. Data-driven multi-scale multi-physics models to derive process-structure-property relationships for additive manufacturing components. *J. Comput.-Aided Mater. Des.* **2018**, *25*, 1–18.