Systems & Control Transactions

Original Research Article
Peer Reviewed Conference Proceeding
ESCAPE 36 - European Symposium on Computer Aided Process Engineering
Sheffield, UK. 21-24 June 2026

PSE PRESS

MatStudio: A Human-in-the-Loop Framework for Microstructure Segmentation with SAM-Guided Refinement

Yao Xue[a], Yanhu Wang[b], Antonios Armaou[cd]*

a The Pennsylvania State University, Department of Mechanical Engineering, University Park, PA 16802, USA
b Wenzhou University, College of Mechanical and Electrical Engineering, Wenzhou, 325035, Zhejiang, China
c University of Patras, Department of Chemical Engineering, Rio, 26504, Greece
d The Pennsylvania State University, Department of Chemical Engineering, University Park, PA 16802, USA
* Corresponding Author: armaou@psu.edu; armaou@upatras.gr

ABSTRACT

Microstructure segmentation is essential for quantitative materials analysis; however, supervised deep learning demands substantial annotation, whereas general-purpose foundation models such as the Segment Anything Model (SAM) offer limited domain-specific semantic control. This paper presents MatStudio, a human-in-the-loop framework for microstructure segmentation that is proposed and implemented end to end in this work. MatStudio couples an interactive workflow for batchwise micrograph annotation and model adaptation with a dual-head convolutional architecture and SAM-guided boundary refinement. The loop combines sparse supervision with SAM-assisted labeling, task-specific training, and iterative batch-level correction, typically converging within two to three cycles.. The network comprises a shared encoder initialized from a pretrained backbone and two decoders: a UNet-style segmentation head that jointly predicts class labels and pixelwise uncertainty, and a prototype branch that measures texture similarity against a memory bank of learnable prototypes. An uncertainty-aware training objective couples label estimation with confidence; uncertainty weighting emphasizes prototype learning in ambiguous regions, and purity filtering restricts prototype updates to phase-interior features to mitigate boundary contamination. At inference, decoder outputs are fused in a spatially adaptive manner according to local confidence. We further introduce SAMFusion, which treats SAM as a morphology-aware proposal source and accepts proposals that are consistent with the fused prediction under a hierarchical small-to-large activation criterion, yielding improved boundary fidelity without case-specific tuning of SAM prompts. Experimental evaluation on Ni-alloy micrographs and public benchmarks shows consistent gains relative to interactive scribble-based segmentation and SAM-centric reference methods.

Keywords: Machine Learning, Artificial Intelligence, Materials, Microstructure segmentation, Human-in-the-loop, Uncertainty quantification, Prototype learning, Segment Anything Model

INTRODUCTION

Microstructure governs the mechanical, thermal, and chemical properties of engineering materials, making quantitative microstructural characterization essential for process optimization and structure-property relationship discovery [1, 2]. The spatial distribution of phases, precipitates, grain boundaries, and defects directly influences macroscopic performance metrics such as strength, ductility, and corrosion resistance. As materials informatics advances toward high-throughput analysis and autonomous experimentation, accurate and efficient microstructure segmentation has become a critical bottleneck.

Traditional image analysis approaches relied on threshold-based methods, morphological operations,

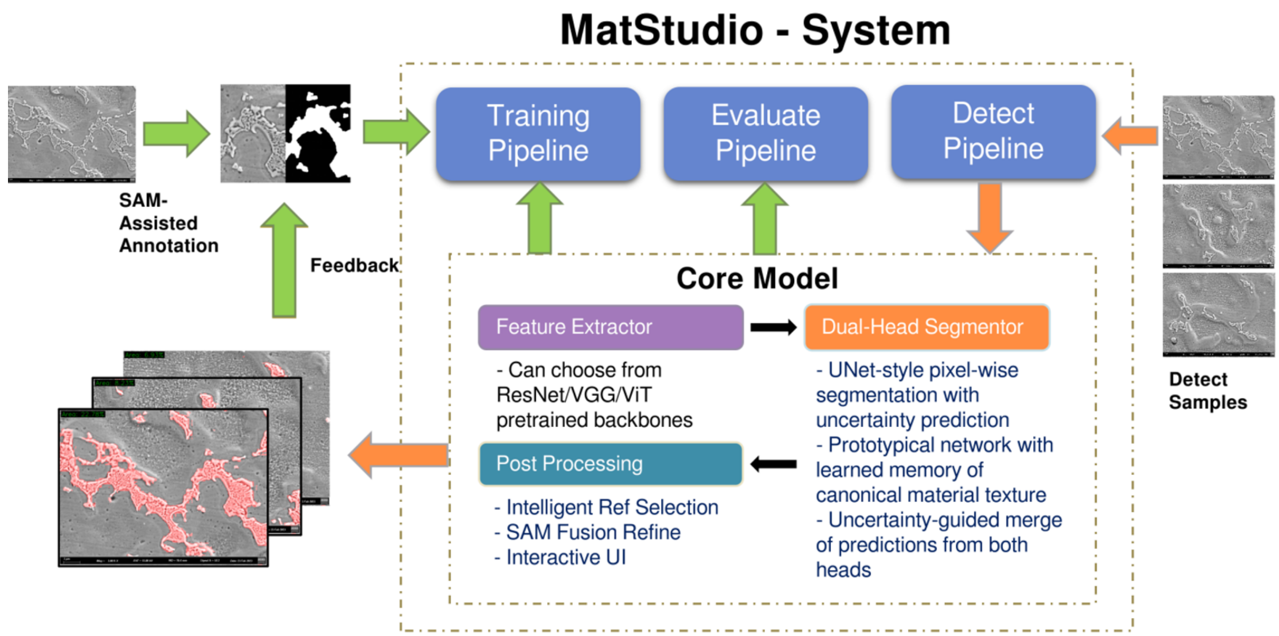

Figure 1. System-level overview of MatStudio. Training, evaluation, and detect interface with a shared core model (backbone, dual-head segmentor, and post-processing); feedback closes the loop between predictions and SAM-assisted annotation.

and hand-crafted features, which require extensive manual parameter tuning and exhibit poor generalization across imaging conditions. The advent of deep learning has substantially advanced automation: U-Net [3] and fully convolutional architectures demonstrated remarkable performance on biomedical and materials images, with subsequent work applying these methods to steel microstructures [4] and precipitate quantification. Pretrained encoders on large microscopy datasets have further improved generalization [5]. However, standard encoder-decoder architectures optimize for pixel-wise classification and may struggle to maintain consistent predictions across texturally homogeneous regions, particularly when training data is limited [6]. This limitation motivates complementary approaches based on metric learning.

Siamese networks [7] address few-shot recognition by learning embeddings where similar samples cluster together, enabling classification via distance comparison to reference examples. Prototypical networks [8] extend this paradigm by representing each class as a single prototype vector—the mean embedding of support examples—and classifying queries based on proximity to these prototypes. For semantic segmentation, PANet [9] demonstrated that prototype alignment enables effective few-shot segmentation by matching query features to class-specific prototypes derived from support masks. These metric-learning approaches excel at capturing texture consistency: once a canonical texture representation is established, regions sharing that texture can be reliably identified regardless of their spatial distribution. However, prototype-based methods typically operate at reduced resolution and may produce coarse boundaries, while their performance depends heavily on the quality and purity of the reference examples used to construct prototypes.

More recently, foundation models trained on massive datasets have demonstrated remarkable zero-shot generalization capabilities. The Segment Anything Model (SAM) [10], trained on over one billion masks, has attracted particular attention for materials microscopy. MatSAM [11] employs automated prompt generation to leverage SAM's zero-shot segmentation for microstructure extraction, SAM-I-Am [12] uses geometric and textural features for semantic boosting in electron microscopy, and SAMRefiner [13] adapts SAM for universal mask refinement. In parallel, Human-in-the-Loop (HITL) frameworks such as scribble-based segmentation [14] have effectively mitigated data scarcity through iterative expert feedback.

Despite these advances, each paradigm exhibits inherent trade-offs. Standard segmentation networks excel at boundary delineation but may produce inconsistent predictions in material interiors where local context alone is insufficient. Prototype-based methods enforce texture consistency but sacrifice spatial precision and require careful reference selection to avoid prototype pollution from boundary features. SAM-based zero-shot approaches achieve impressive results without training, but directly applying SAM to material microscopy poses challenges due to the dense and dispersed nature of microstructural features; moreover, these methods create dependency on prompt quality where suboptimal prompts directly degrade results. This analysis suggests an opportunity: rather than relying on any single approach, robust microstructure segmentation may be achieved by combining texture-consistent prototype matching with pixel-precise segmentation, while leveraging SAM's morphological sensitivity for boundary refinement.

Building on this insight, we present MatStudio, a unified framework that integrates an uncertainty-guided dual-head architecture with SAMFusion—a novel proposal-guided refinement strategy. The primary segmentation head produces pixel-wise predictions along with learned uncertainty estimates [15], which guide both training and inference of an auxiliary prototype memory bank. During training, high-uncertainty regions receive amplified supervision for the prototype branch, forcing it to specialize on cases where the segmentation head struggles; at inference, predictions are dynamically fused based on local confidence. SAMFusion further refines boundaries by treating SAM purely as a morphology-aware proposal generator, activating proposals only when corroborated by the fused prediction. This design inherits the boundary precision of segmentation networks, the texture consistency of prototype matching, and SAM's morphological sensitivity—while avoiding the failure modes of each approach used in isolation.

The remainder of this article is organized as follows. Section 2 presents the MatStudio methodology. Figure 1 gives a system-level overview; the text then details the interactive workflow, the dual-head architecture with prototype-based texture modeling, uncertainty-guided training and memory updates, spatially adaptive fusion at inference, and SAMFusion for proposal-gated boundary refinement. Section 3 reports comparative experiments on Ni-alloy micrographs and public benchmarks against ScribbleSeg and MatSAM, including the qualitative assessment in Figure 2. Section 4 summarizes the principal findings and discusses implications for interactive microstructure segmentation.

METHODOLOGY

Figure 1 sketches MatStudio at architecture level: training, evaluation, and detect pipelines connect to a shared core model, with outputs feeding back to annotation. The operational workflow follows a human-in-the-loop paradigm for narrow batches of micrographs under consistent imaging conditions. Users begin by loading a batch into an interactive annotation interface and providing sparse labels—point clicks or bounding boxes—for the target microstructural features. These annotations, optionally accelerated by SAM-assisted region proposals, bootstrap a task-specific segmentation model. After an initial training pass, the model generates predictions across the batch; the user then reviews these outputs and provides corrective feedback on under-segmented or mis-segmented regions, which refines the supervision signal for the next iteration. This iterative refinement cycle continues until the segmentation quality converges, typically within two to three rounds.

Dual-Head Architecture with Prototype Memory

The core segmentation model employs a dual-head architecture where a primary segmentation decoder is complemented by an auxiliary prototype-based similarity head. A shared convolutional encoder, initialized from a pretrained backbone, extracts hierarchical feature representations that feed two parallel branches: (1) a UNet-style segmentation head producing dense pixel-wise predictions along with a learned uncertainty signal, and (2) a prototype head that computes texture similarity against a dynamically maintained memory bank of canonical material textures.

The segmentation head is designed to output not only class predictions but also a per-pixel confidence estimate, inspired by heteroscedastic uncertainty modeling in Bayesian deep learning [15]. This enables the network to explicitly signal regions where it lacks confidence—typically phase boundaries, texturally ambiguous zones, and imaging artifacts. The auxiliary prototype head maintains a compact set of learnable prototype vectors representing characteristic material textures. For each spatial location, the head performs texture similarity matching against all prototypes and outputs a dense similarity map that captures texture consistency across the image.

2.2 Uncertainty-Guided Learning

The key insight of our training strategy is that the segmentation head's uncertainty provides a natural supervisory signal for the auxiliary branch. The segmentation head is trained with an uncertainty-aware loss formulation that jointly learns predictions and confidence: when prediction error is high, the network learns to increase its uncertainty estimate, effectively admitting difficult regions. At equilibrium, the uncertainty map captures precisely those regions where the segmentation head consistently struggles.

This learned uncertainty is then used to modulate supervision for the prototype head. Regions with high uncertainty receive amplified training signal, directing the prototype branch to specialize on cases the segmentation head cannot handle reliably. This uncertaintyweighted training ensures complementary specialization: the segmentation head handles straightforward regions while the prototype head focuses its learning capacity on challenging textures and ambiguous boundaries. The total training objective combines both components with appropriate weighting.

To maintain prototype quality, we employ a conservative update strategy with strict purity filtering. Before updating prototypes, we identify regions that are unambiguously inside material phases—far from boundaries—and use only these "pure" features to update the memory bank. This prevents prototype pollution from boundary features, which would otherwise cause boundary blurring and increased false positives.

Inference and SAMFusion Refinement

At inference time, the learned uncertainty directly governs how predictions from both heads are combined. In confident regions, the segmentation head dominates; in uncertain regions, the prototype-based similarity takes over. This dynamic, spatially-adaptive fusion requires no additional hyperparameters and automatically adapts to local image content.

To further enhance boundary quality, we introduce SAMFusion, a proposal-guided refinement strategy that leverages SAM's exceptional morphological sensitivity without inheriting its failure modes. Unlike approaches that treat SAM outputs as authoritative segmentations—and consequently suffer when SAM misinterprets complex microstructures—SAMFusion treats SAM purely as a morphology-aware proposal generator. We invoke SAM to obtain an exhaustive set of candidate regions, then selectively activate proposals through an agreement-gated mechanism: proposals are accepted only when they exhibit sufficient overlap with the network's fused prediction. This hierarchical activation process, which evaluates proposals from small to large, ensures that SAM's boundary precision refines the segmentation without introducing spurious detections in texturally ambiguous regions. Crucially, this design eliminates the need for case-by-case hyperparameter tuning of SAM's prompt strategy.

RESULTS

We benchmarked MatStudio against ScribbleSeg [14] and MatSAM [11] on four segmentation tasks spanning self-collected Ni-alloy micrographs (cases a–b) and public datasets—UHCS spheroidite [16] and NBS-2 eutectic carbides [14]—as shown in Figure 2.

Cases (a) and (b) expose complementary failure modes of the baselines. In (a), dendritic arms spanning the full field-of-view defeat MatSAM's edge-based prompting, which cannot reliably distinguish foreground from background when boundaries lack clear termination; ScribbleSeg under-segments due to textural ambiguity between dendrites and intermediate grey substructures in the matrix. In (b), off-axis detector gradients and secondary-electron edge halos generate spurious boundaries that MatSAM interprets as phase interfaces, producing extensive false positives; the same gradient also causes ScribbleSeg to fail.

Case (c) presents well-defined spheroidite particles that all methods largely detect; however, surface cracks partition the matrix into coarse enclosed domains that, despite containing no target particles, exhibit closed-contour morphology indistinguishable from valid regions under edge-based prompting, causing MatSAM to register these large spurious areas as false positives. Case (d) constitutes the most demanding scenario: eutectic carbides with amorphous, interconnected morphologies cause catastrophic failure in MatSAM regardless of hyperparameter tuning, while ScribbleSeg partially succeeds but cannot resolve regions where carbide and matrix textures overlap. We note that published metrics for ScribbleSeg and MatSAM on these benchmarks could not be reproduced under our experimental protocol; moreover, manual inspection revealed annotation inconsistencies in the ground-truth masks—several target particles remain unlabeled while certain background regions are erroneously marked—rendering absolute metric comparisons unreliable. Despite these caveats, IoU and Dice reported as overlay annotations on each prediction panel in Figure 2 still permit relative comparison among methods on the same exemplars and under our evaluation settings.

The IoU and Dice coefficients printed on those panels summarize performance on the displayed crops. MatStudio attains the highest IoU and Dice in every row: in (a), 0.92 and 0.96 versus 0.44 / 0.62 (ScribbleSeg) and 0.41 / 0.58 (MatSAM); in (b), 0.74 / 0.85 versus 0.15 / 0.26 and 0.56 / 0.72; in (c), 0.66 / 0.80 versus 0.48 / 0.65 and 0.57 / 0.72; in (d), 0.95 / 0.97 versus 0.64 / 0.78 and 0.05 / 0.10. These rankings are consistent with the qualitative failure modes discussed above and should be read together with the protocol and labeling limitations noted earlier.

Across all four cases, and in agreement with both the visual evidence and the panel-reported scores in Figure 2, MatStudio maintains robust performance by exploiting the complementarity between learned texture representations and SAM's morphological sensitivity. The agreement-gated activation mechanism filters out proposals that, while geometrically plausible, lack corroboration from the semantic predictor—effectively suppressing edge artifacts and spurious enclosed regions. Crucially, by treating SAM outputs as candidate refinements rather than authoritative segmentations, the framework inherits SAM's boundary-delineation strengths without suffering its failure modes when edge-

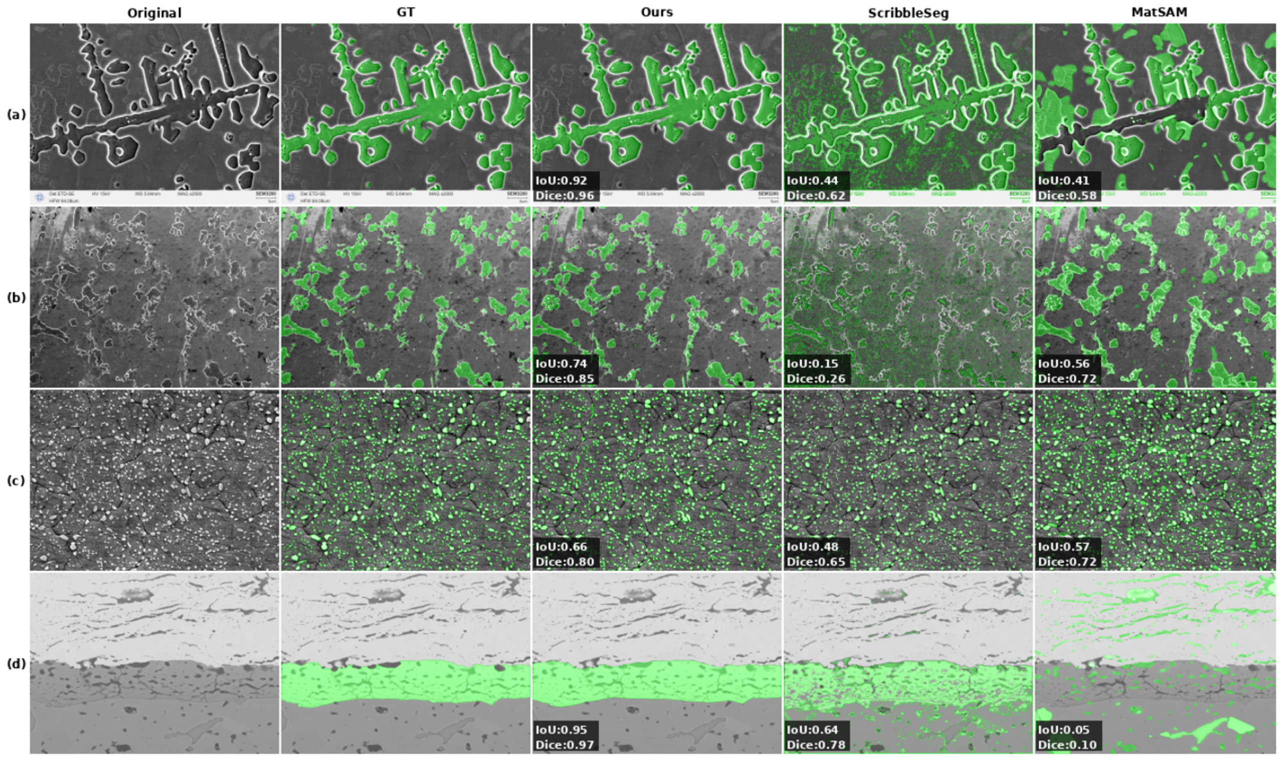

Figure 2. Qualitative comparison across four material systems. Columns show original image, ground truth, and predictions from MatStudio (ours), ScribbleSeg, and MatSAM. IoU and Dice scores are presented in overlay annotations. We observe that scribble-seg fails when there is no clear textural difference between the target and background. Similarly, MatSAM fails when boundary of the target area is complex or arbitrary. Our MatStudio is capable of delivering consistent and robust performance across difference scenarios and outperforms its counterparts in most cases.

based prompting encounters ambiguous or artifact-laden microstructures.

CONCLUSION

We presented MatStudio, a human-in-the-loop framework for microstructure segmentation that combines an uncertainty-guided dual-head model with SAMFusion for proposal-gated boundary refinement. Across the heterogeneous microscopy settings considered in this work, the comparison in results section illustrates a recurring pattern: methods that rely predominantly on edge- or prompt-driven segmentation exhibit substantial fragility when microstructural morphology, contrast, or imaging artifacts violate the assumptions those pipelines implicitly encode, whereas MatStudio remains comparatively stable by coupling texture-aware predictions with conservative use of SAM as a geometric proposal source. The dual-head design uses estimated uncertainty to allocate responsibility between dense segmentation and prototype-driven texture reasoning at inference, and SAMFusion retains SAM refinements only when they agree with the fused model, reducing spurious activations that arise when proposals are treated as definitive masks. We report exemplar-level overlays under our training and evaluation protocol together with the limitations discussed in the Results section; accordingly, we frame the findings as consistent relative evidence for the proposed integration rather than as universal rankings or as claims of proportional improvement over prior published numbers.

AUTHOR IDENTIFIERS

Author ORCIDs:
Xue Y: 0000-0002-2944-0005
Wang Y: 0009-0001-1200-6157
Armaou A: 0000-0002-8592-7934

REFERENCES

1. Cecen A, Dai H, Yabansu YC, Kalidindi SR, Song . Material structure-property linkages using three-dimensional convolutional neural networks. Acta Materialia 146:76-84 (2018).

https://doi.org/10.1016/j.actamat.2017.11.053
2. Shen C, Wang C, Wei X, Li Y, van der Zwaag S, Xu W. Physical metallurgy-guided machine learning and artificial neural network for high-strength alloy design. Acta Materialia 2021;205:116601.
3. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science :234-241 (2015). https://doi.org/10.1007/978-3-319-24574-4_28
4. Azimi SM, Britz D, Engstler M, Fritz M, Mücklich F. Advanced steel microstructural classification by deep learning methods. Sci Rep 8: (2018). https://doi.org/10.1038/s41598-018-20037-5
5. Stuckner J, Harder B, Smith TM. Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset. npj Comput Mater 8: (2022). https://doi.org/10.1038/s41524-022-00878-5
6. Yang L, Zhang Y, Chen J, Zhang S, Chen DZ. Suggestive annotation: a deep active learning framework for biomedical image segmentation. Lecture Notes in Computer Science :399-407 (2017). https://doi.org/10.1007/978-3-319-66179-7_46
7. Koch G, Zemel R, Salakhutdinov R. Siamese Neural Networks for One-shot Image Recognition. In: ICML Deep Learning Workshop 2015.
8. Snell J, Swersky K, Zemel R. Prototypical Networks for Few-shot Learning. In: NeurIPS 2017. p. 4077-4087.
9. Wang K, Liew JH, Zou Y, Zhou D, Feng J. Panet: few-shot image semantic segmentation with prototype alignment. 2019 IEEE/CVF International Conference on Computer Vision (ICCV) :9196-9205 (2019). https://doi.org/10.1109/iccv.2019.00929
10. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg AC, Lo WY, Dollár P, Girshick R. Segment anything. 2023 IEEE/CVF International Conference on Computer Vision (ICCV) :3992-4003 (2023). https://doi.org/10.1109/iccv51070.2023.00371
11. Li C, Han X, Yao C, Guo Y, Li Z, Jiang L, Liu W, Huang H, Fu H, Ban X. A novel training-free approach to efficiently extracting material microstructures via visual large model. Acta Materialia 290:120962 (2025). https://doi.org/10.1016/j.actamat.2025.120962
12. Abebe W, Strube J, Guo L, Tallent NR, Bel O, Spurgeon S, Doty C, Jannesari A. Sam-i-am: semantic boosting for zero-shot atomic-scale electron micrograph segmentation. Computational Materials Science 246:113400 (2025). https://doi.org/10.1016/j.commatsci.2024.113400
13. Deng X, Wu H, Zeng R, Qin J. Memsam: taming segment anything model for echocardiography video segmentation. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) :9622-9631 (2024). https://doi.org/10.1109/cvpr52733.2024.00919
14. Na J, Kim SJ, Kim H, Kang SH, Lee S. A unified microstructure segmentation approach via human-in-the-loop machine learning. Acta Materialia 255:119086 (2023). https://doi.org/10.1016/j.actamat.2023.119086
15. Kendall A, Gal Y. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? In: NeurIPS 2017. p. 5574-5584.
16. DeCost BL, Lei B, Francis T, Holm EA. High throughput quantitative metallography for complex microstructures using deep learning: a case study in ultrahigh carbon steel. Microsc Microanal 25:21-29 (2019). https://doi.org/10.1017/s1431927618015635

© 2026 by the authors. Licensed to PSEcommunity.org and PSE Press. This is an open access article under the creative commons CC-BY-SA licensing terms. Credit must be given to creator and ad-aptations must be shared under the same terms. See https://creativecommons.org/licenses/by-sa/4.0/