nature portfolio

Peer Review File

Synthetic Data-Driven Deep Learning for Label-free Autonomous Atomic Force Microscopy

Corresponding Author: Dr Ruben Millan-Solsona

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Paper presents a data simulation tool suitable for machine learning training in autonomous bio-AFM. Works like this are very important as they enable building more robust data processing algorithms, all possible types, including algorithms based on use of machine learning, so it is significant for a broad area of AFM data processing tasks.

Paper is nicely written, methodology is sound and results are impressive, so I have only few minor comments:

- It would be much better to talk about bio-AFM specific artefacts instead of AFM specific artefacts as AFM is broadly used also on other areas where samples are very different and the presented methodology for height generation works only for small objects on flat background.

- It would be nice to comment more thermal/mechanical drift impact on the data as this is frequently found source of experimental data distortion and is somehow touched in the introduction but seems not to be developed further in the presented figures.

- Fig. 2e and 2f (and some of figures in 2h) look like line correction was applied which is strange. Line flattening artifacts are mentioned also in the description of the data synthesis. Line flattening is however not part of AFM measurement process but more an effect of postprocessing, so I would guess that data coming from the microscope (real or simulated) should not have such effect and I am not sure if data prepared for training ML models should contain it. This could be further commented, as mixing the effects related to the microscope and related to the postprocessing is a bit confusing.

(Remarks on code availability)

Reviewer #2

(Remarks to the Author)
This manuscript presents SimuScan, a Python-based synthetic data generator designed to produce physically realistic AFM-like images for training deep learning models without manual labeling. The authors demonstrate that models trained exclusively on SimuScan-generated datasets can generalize to experimental AFM data, supporting real-time feature recognition and enabling semi-autonomous AFM imaging workflows. Applications are shown for geometric nanostructures, DNA, and bacterial cells.
The topic is timely, combining synthetic data, AI, and AFM automation. The manuscript is clearly written and technically sophisticated. However, several methodological and validation shortcomings weaken the central claim of generalization from synthetic to real data. I view this work as potentially impactful but requiring significant additional evidence and clarification before publication.

Major Comments
The manuscript asserts that models trained exclusively on synthetic data generalized effectively to experimental AFM images. However, the presented validation relies on very limited experimental datasets (25 manually labeled images per

morphology, expanded by augmentation).
• Such small datasets cannot robustly support claims of cross-domain generalization.
• Performance metrics in Table 1 appear to be computed primarily on synthetic test sets.
• The evaluation on real data (DNA, bacteria) is qualitative and anecdotal.
The introduction and results sections are well written but somewhat repetitive and promotional, particularly around AI/AFM "revolution" claims.
• Statements like "redefining what is achievable" and "establishes a foundation for self-guided experimentation" could be toned down.
• The paper would benefit from clearer differentiation between demonstrated and aspirational capabilities.

Personally, I would recommend including quantitative evaluation on independent experimental datasets, with metrics such as IoU, mAP@[0.5:0.95], and F1 scores across several morphologies.
While the authors describe SimuScan as unique, no comparisons are made with existing synthetic AFM data generators (e.g., BioAFMviewer, Degenhardt et al.). Without such benchmarks, it is unclear whether SimuScan provides superior realism or utility for training.
Adding a benchmark against at least one existing synthetic dataset or model trained on small real datasets to quantify the improvement due to SimuScan.

More specifically about the Figures and Tables, I would like to comment them:
Figure 1 – Strong conceptual overview but overly abstract. Incorporating example AFM images or segmentation overlays and color-coding synthetic versus experimental stages would improve clarity.
Figure 2 – Excellent depiction of the SimuScan pipeline but visually overloaded. Splitting into two figures—(1) pipeline schematic and (2) dataset examples—would aid comprehension. Quantitative validation insets comparing real and synthetic textures should be added.
Table 1 – Informative yet difficult to interpret. Reformat as a clear matrix with error bars or replace colored dots with shaded highlights for print clarity. A complementary bar chart could visualize performance differences.
Figure 3 – The closed-loop workflow schematic is intuitive but lacks experimental imagery. Numbered process steps and inclusion of a real scan inset would make it more instructive.
Figure 4 – Impressive quantitative analysis of P. aeruginosa cells. Add scale bars, representative zoom-ins, and explicit sample size (n = 100) with labeled units on plots. Histogram fits should list model parameters.
Figure 5 – The large-area mapping of mixed bacterial populations is visually striking but needs validation of bacterial identity (e.g., optical overlay) and explicit scales for density maps and insets.
Across all figures, maintain uniform color scales, include scale bars, and use accessible color palettes for color-blind readers (just a suggestion)

The "autonomous AFM" section presents a compelling integration of real-time segmentation and stage control, yet the process still involves human-defined parameters, thresholds, and stopping criteria. The absence of discussion of failure modes (e.g., misdetections, tip crashes, or drift correction) overstates the level of autonomy. The SPM community wants to know!
By recasting this section as a semi-autonomous or guided AFM workflow, and explicitly describing system limitations and failure recovery strategies.
Although SimuScan includes key AFM-specific artifacts (tip convolution, noise, flattening), their quantitative accuracy is not evaluated. It is unclear whether the simulated distortions statistically resemble real AFM data. This point is also important to know for the end-users.
Providing a quantitative comparison between simulated and experimental AFM images, for instance, using height–height correlation functions, noise spectra, or power spectral density analysis will certainly be helpful.
The paper states that SimuScan is "available upon request," which contradicts the claim of open science. Reproducibility depends on full code and configuration availability.
My recommendations will be to publicly release the full SimuScan source code and documentation (or a Docker container) to ensure replicability, even though some parts are on GitHub and I haven't tested it so far … but I certainly will!
The bacterial and DNA demonstrations are visually appealing but lack quantitative biological validation. The models' specificity (ability to distinguish true cells from debris) and robustness to sample variation remain unclear.
Maybe the authors could include, in SI section, confusion matrices or precision/recall data for biological classifications, ideally compared to manual expert annotation.

Minor Comments
1. Terminology: In Figure 4, "Bacillus cells" likely refers to Pseudomonas aeruginosa; please correct to avoid taxonomic confusion.
2. Metrics: Use consistent metrics (preferably mAP@[0.5:0.95]) across models for fair comparison.
3. Figures: Figures 2–5 contain excessive subpanels and dense captions; consider streamlining for readability.
4. Data Volume: Clarify how 5,000 synthetic images suffice for training deep architectures—were data augmentations or tiling used?
5. Manuscript Tone: Several statements (e.g., "redefining what is achievable") are overly promotional; suggest tempering to reflect demonstrated results.
6. Formatting: Minor typographical inconsistencies (μ vs μ, superscripts, spacing) should be corrected.
In conclusion, this work presents an exciting and potentially field-shifting approach to autonomous AFM. However, its claims of synthetic-to-real generalization and full autonomy require stronger experimental evidence, quantitative benchmarking, and improved reproducibility. With these revisions, the paper could make a substantial contribution to autonomous nanometrology and AI-driven instrumentation.

I recommend major revision but I strongly believe that by addressing my comments and remarks, the paper will be among the key-ones in this growing field merging SPM and AI.

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
Millan-Solsona et al present a novel tool, SimuScan, to create artificial AFM datasets based on specific criteria. They show that autonomous operation of AFM based on artificially-trained machine learning (ML) is possible over a range of samples, from simple geometrical patterns on hard substrates to bacteria. They also emphasise the analysis capabilities of ML tools trained on with SimuScan, as illustrated on DNA structures.
The authors argue that the absence of tractable and consistent datasets for ML is the largest bottleneck to wide adoption of ML strategies in AFM. I agree with the authors about this being a major roadblock, both for automation and image interpretation. Several attempts to create extensive AFM databases for ML have not succeeded for lack of suitable large sets and/or metadata information. Most successful publications tend to focus on a narrow and well defined problem.
In this context, SimuScan clearly offers a step in the right direction. However, it remains to be seen whether this step is incremental or a step-change.

My main concern with SimuScan as presented here is that it is illustrated on clearly defined and easily identified criteria for samples such as geometrical shapes or loop vs elongated strand. While likely sufficient for large (>100 nm) samples, I am not convinced that it is enough for nanoscale images where countless known and unknown effects may be at play, from convolution to intermittent molecular contamination of the tip, asymmetric tip shapes, unstable sample fixation, dirty sample, inappropriate scanning parameters, etc. I feel this is important because the power of AFM lies at the nanoscale, where optical approaches are less straightforward. The authors show that SimuScan can distinguish different DNA shapes, but this is still a relatively simple question regardless of the possible issues mentioned above. What about distinguishing the orientation of an adsorbed (small) molecule? Or a clean from a dirty tip? Generally, how can realistic nanoscale data be generated without knowing all the details about the sample and the tip-sample interactions that can dramatically change the resulting image? Is this a limitation of SimuScan?

A second concern/question is the apparent lack of spectroscopic information. Currently, most AFM operate in spectroscopy mode (peak force tapping, wave mode, etc) with considerable information in the spectroscopy curve obtained at each pixel. Can SimuScan create relevant information to train ML of the spectroscopy curves?

The presentation of the paper is generally clear, but Fig 2 contains too much information (and too small in h/i) to be easily appreciated.

(Remarks on code availability)

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
All my questions were answered and with large help of questions of the other referees the manuscript was significantly improved. From my point of view it is now acceptable for publication.

(Remarks on code availability)
The code seems to contain everything necessary and seems to be logically structured, but I did not try to run it or test it in detail.

Reviewer #2

(Remarks to the Author)
Nice improvement according to the comments and suggestions made by the three referees!

(Remarks on code availability)

Reviewer #3

(Remarks to the Author)
The authors have extensively revised their manuscript and addressed the bulk of my concerns. I am happy to recommend publication after they have revised typos and textual mistake (e.g. references mis-formatted: "Stage automation extends this

process beyond the intrinsic ~100 µm AFM scan range,{Millan-Solsona, 2025 #44} enabling tiling across large sample areas and revisiting of features identified by segmentation".)

(Remarks on code availability)

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Paper presents a data simulation tool suitable for machine learning training in autonomous bio-AFM. Works like this are very important as they enable building more robust data processing algorithms, all possible types, including algorithms based on use of machine learning, so it is significant for a broad area of AFM data processing tasks.

Paper is nicely written, methodology is sound and results are impressive, so I have only few minor comments:

We sincerely thank the reviewer for the positive assessment of our manuscript, and for recognizing the significance of physics-informed simulation tools for enabling robust AFM data processing and machine-learning–based automation.

We appreciate the reviewer's comments regarding the quality of the writing, methodological soundness, and strength of the results. Below, we respond point-by-point to the minor comments and describe the corresponding revisions, including clarifications that further emphasize SimuScan's broad applicability across AFM modalities and sample classes, as well as additional analyses/datasets added in the revised manuscript.

- It would be much better to talk about bio-AFM specific artefacts instead of AFM specific artefacts as AFM is broadly used also on other areas where samples are very different and the presented methodology for height generation works only for small objects on flat background.

We thank the reviewer for raising this important point and for the opportunity to clarify the intended scope of the framework.

A central motivation for SimuScan is that, relative to optical and electron microscopy, AFM lacks large, diverse, publicly available annotated datasets. Moreover, pretrained models from natural-image domains transfer poorly to AFM due to its distinct contrast mechanisms, imaging physics, and artifact structure. Addressing this broader gap, rather than targeting a single sample class or application, was a primary design goal of SimuScan.

While examples in the original submission emphasized biologically relevant samples, this was partly related to project funding, however, we now clarify and demonstrate that SimuScan is designed as a **general AFM framework**, not a bio-AFM–specific tool or workflow. In the revised manuscript, we have expanded the dataset to include nanostructures and more complex, heterogeneous substrates to reflect this broader scope. The goal of this work is to show that SimuScan generalizes across a range of morphologies and enables robust, rapid image segmentation and feature finding, representing a significant step forward for AFM automation. We want to highlight, that in the revised manuscript we introduced self-seeded experimental geometry workflow which allows the synthetic data generation to be based on CAD designs or AFM images

themselves, in this was we are not limited to simple objects—biological or inorganic—that can be easily parameterized.

arbitrary experimental AFM morphologies—biological or inorganic—to be converted into STL representations and used to generate large, morphology-preserving synthetic datasets. This removes assumptions about object size, background flatness, or sample class, further broadening applicability

The synthetic image generation focuses on the topographic channel and incorporates artifact mechanisms—such as tip–sample convolution, electronic and 1/f noise, drift, and feedback- and line-related distortions—that are intrinsic to AFM operation and occur across both materials and biological sciences. For task-specific problems (e.g., in situ heating, dynamic biological systems) or samples requiring highly refined, application-dependent models, SimuScan is intended to serve as a robust starting point. In such cases, optimal performance can be achieved by tuning the synthetic data generation parameters and model hyperparameters within the SimuScan workflow.

Importantly, the framework is **not restricted to small objects on flat backgrounds**. As demonstrated in the revised manuscript (see Figure S3 and S8) , SimuScan explicitly incorporates substrate topography during synthetic data generation. To validate this capability, we investigated non-planar, including samples with debris/contamination as well as structured substrates wherein bacterial cells were deposited on nanopatterned SiNx pillar arrays. In this latter scenario, models trained solely on planar substrates failed to reliably detect cells. In contrast, as shown in Fig. S8, incorporating substrate morphology within the SimuScan workflow enabled robust segmentation on such topographically complex surfaces. Ongoing experiments indicate that this behavior generalizes across a variety of substrates and surface contaminations commonly encountered in AFM.

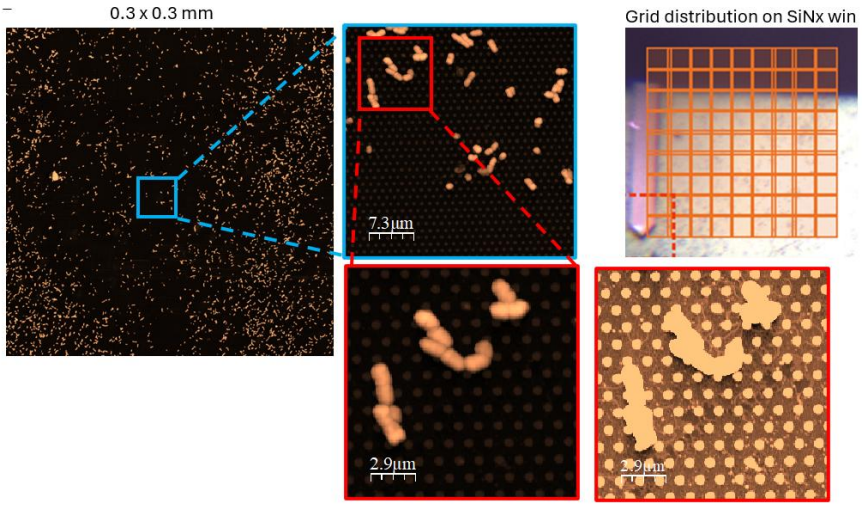

Figure 1. (Left) Large-area AFM topography (0.3 × 0.3 mm) showing sparse bacterial coverage on a nanopatterned SiNx membrane; the dashed box marks a region selected for higher-resolution imaging. (Center) Intermediate- and high-magnification views reveal individual bacterial cells interacting with the underlying pillar lattice. (Right) Optical image of the SiNx

membrane illustrating the predefined grid layout for systematic large-area scanning, alongside a saturated AFM image highlighting the substrate topography. These samples challenge models trained on planar substrates; however, since its possible to tune SimuScan's synthetic training data to account for substrate-specific effects, accurate cell–substrate segmentation is achieved (Fig. S8).

To avoid confusion, we have revised the manuscript to (i) consistently refer to **AFM-specific**, rather than bio-AFM–specific, artifacts; (ii) explicitly emphasize applicability across both biological and non-biological material systems; and (iii) expand discussion and examples demonstrating performance on structured and heterogeneous substrates. Together, these changes clarify that SimuScan is intended as a **general AFM data-generation and automation framework**, not a tool limited to bio-AFM scenarios.

- It would be nice to comment more thermal/mechanical drift impact on the data as this is frequently found source of experimental data distortion and is somehow touched in the introduction but seems not to be developed further in the presented figures.

Response:
We thank the reviewer for highlighting the importance of thermal and mechanical drift, which is indeed a common source of distortion in AFM experiments. SimuScan includes the capability to simulate gradual displacements and distortions that mimic drift effects, with adjustable magnitude and direction. However, to maintain focus on the artifact mechanisms most critical for deep-learning training in the present study—namely noise, tip–sample convolution, and line-related distortions—we chose not to explicitly include drift simulations in the main figures.

All experimental AFM data presented in this work were acquired using closed-loop scanners, and at the scan sizes and acquisition times considered here, drift effects were minimal. Any residual drift was further mitigated using standard line-alignment and flattening procedures, resulting in negligible drift in the datasets used for evaluation.

To clarify the scope of the framework, we have added a statement in the revised manuscript noting that using the Simuscan platform it is possible to include incorporate and tune drift levels relevant for specific experimental scenarios. While the current implementation intentionally emphasizes the topographic channel and the most prevalent AFM artifacts for segmentation and feature identification, extending the framework to include long-timescale drift dynamics, additional contrast channels, and time-dependent probe effects represents a natural and modular direction for future development. These extensions would further expand the realism and applicability of SimuScan across a wider range of AFM modalities and experimental conditions.

While SimuScan captures a broad range of AFM-specific artifacts relevant for realistic synthetic data generation, this first implementation makes deliberate scope choices leaving clear avenues for future extension. At present, the framework focuses on the topographic channel, which is broadly applicable across common AFM modes (e.g., dynamic, static, PeakForce Tapping, and wave mode -based operation), while additional contrast channels such as phase, amplitude, or

spectroscopic signals are not explicitly modeled. Long-timescale thermal and mechanical drift is also not simulated, reflecting the prevalence of closed-loop AFM systems and standard post-processing procedures that mitigate these effects in many experiments. In addition, SimuScan models static double-tip artifacts but does not yet capture intermittent or time-dependent probe contamination. Extending the framework to incorporate additional signal channels, drift dynamics, and evolving probe conditions represents a natural and modular path forward, and would further expand the realism, versatility, and applicability of SimuScan across a wider range of AFM modalities and experimental conditions.

- Fig. 2e and 2f (and some of figures in 2h) look like line correction was applied which is strange. Line flattening artifacts are mentioned also in the description of the data synthesis. Line flattening is however not part of AFM measurement process but more an effect of postprocessing, so I would guess that data coming from the microscope (real or simulated) should not have such effect and I am not sure if data prepared for training ML models should contain it. This could be further commented, as mixing the effects related to the microscope and related to the postprocessing is a bit confusing.

Response:

We thank the reviewer for this thoughtful comment and agree that it is important to clearly distinguish between artifacts arising from AFM acquisition and those introduced during post-processing. We have revised the manuscript to explicitly clarify this distinction.

Although line flattening is not part of raw AFM acquisition, it is a ubiquitous post-processing step in experimental practice, routinely applied to correct residual drift, scanner tilt, and contrast variations. Because flattening is commonly applied prior to analysis and automation, it can introduce secondary artifacts such as residual line features or local distortions. SimuScan therefore optionally incorporates flattening-induced artifacts to improve the realism of synthetic datasets and to enhance model robustness to post-processing effects encountered in practical AFM workflows.

Importantly, SimuScan provides full user control over whether flattening is applied, as well as over the type and magnitude of the correction. In the workflow presented here, flattening was applied prior to object detection to reflect common practice in automated AFM pipelines. This clarification has been added to the manuscript to avoid ambiguity regarding the origin and role of line-related artifacts.

Finally, post-processing effects, including line-flattening errors, are applied to replicate distortions commonly observed in experimental AFM datasets. Although flattening is not part of raw AFM acquisition, it is routinely applied to correct drift, scanner tilt, and contrast variations and can introduce secondary artifacts such as residual line features or local distortions. SimuScan explicitly incorporates flattening-induced artifacts to improve dataset realism and enhance model

robustness to post-processing effects encountered in practice. Users retain control over whether flattening is applied, as well as over the type and magnitude of the correction; in this workflow, flattening is applied prior to object detection to reflect common automated AFM practice.

Reviewer #2 (Remarks to the Author):

This manuscript presents SimuScan, a Python-based synthetic data generator designed to produce physically realistic AFM-like images for training deep learning models without manual labeling. The authors demonstrate that models trained exclusively on SimuScan-generated datasets can generalize to experimental AFM data, supporting real-time feature recognition and enabling semi-autonomous AFM imaging workflows. Applications are shown for geometric nanostructures, DNA, and bacterial cells.
The topic is timely, combining synthetic data, AI, and AFM automation. The manuscript is clearly written and technically sophisticated. However, several methodological and validation shortcomings weaken the central claim of generalization from synthetic to real data. I view this work as potentially impactful but requiring significant additional evidence and clarification before publication.

Response:
We thank the reviewer for their careful and constructive evaluation of the manuscript and for recognizing both the timeliness and technical sophistication of integrating synthetic data, AI, and AFM automation. We agree that demonstrating robust generalization from synthetic to experimental AFM data is central to the contribution and must be supported by rigorous validation.

In response to the reviewer's concerns, we have undertaken substantial new work that significantly strengthens both the methodological foundation and the experimental evidence. The revised manuscript now includes:

I. expanded quantitative benchmarking evaluated **exclusively on independent experimental AFM datasets**;
II. new demonstrations spanning **diverse morphologies**, including biological assemblies, engineered nanostructures, and non-planar substrates;
III. and **algorithmic extensions** that broaden the physical realism and applicability of SimuScan.

Collectively, these additions directly address the question of generalization and substantially reinforce the central claims of the paper.

Most notably, we introduce two complementary, non-parameterized geometry pathways designed explicitly to overcome limitations associated with morphology-specific modeling. We present a **self-seeded experimental geometry workflow**, in which a single experimentally acquired AFM image is converted into a three-dimensional STL surface that serves as a seed morphology for generating large synthetic training datasets. This approach moves beyond the purely parameterized framework of the original submission and enables SimuScan to adapt directly to arbitrary experimental morphologies, while systematically introducing controlled variability in shape, texture, deformation, noise, tip–sample interactions, and substrate context. This mechanism provides a principled and scalable route to synthetic-to-experimental transfer across heterogeneous AFM datasets.

Finally, we emphasize that the goal of this work is **not to exhaustively optimize segmentation models for a specific task**—an effort that typically requires extensive manual annotation,

architecture tuning, and task-specific hyperparameter optimization. Rather, the primary contribution of SimuScan is to enable **rapid, reliable generation of physically grounded, pre-annotated AFM datasets** that substantially lower the barrier to training effective models. In the attached manuscript we show that this approach is robust and generalizable, however, we also make a point to stress that for applications requiring near-perfect precision or highly specialized performance, we fully acknowledge that further task-specific refinement may be necessary. SimuScan is intended to serve as a general, extensible foundation upon which such refinements can be efficiently built.

Below, we respond point-by-point to the reviewer's specific comments and detail the corresponding revisions and new results added to the manuscript.

Major Comments

The manuscript asserts that models trained exclusively on synthetic data generalized effectively to experimental AFM images. However, the presented validation relies on very limited experimental datasets (25 manually labeled images per morphology, expanded by augmentation).

We thank the reviewer for this detailed and constructive critique. We agree that robust, quantitative validation on independent experimental data is essential for substantiating synthetic-to-experimental generalization, and we have revised the manuscript extensively to address each of the points raised.

• Such small datasets cannot robustly support claims of cross-domain generalization.

We thank the reviewer for this comment and agree that claims of cross-domain generalization must be supported by careful validation. We would like to clarify an important point regarding the experimental dataset used in Table 1, as it directly addresses this concern.

As shown in Table 1, the experimental dataset of 25 manually labeled AFM images per morphology is **not used to validate SimuScan-trained models in isolation**, but rather to provide a **direct performance benchmark** against a conventionally trained model. Specifically, we compare (i) models trained exclusively on SimuScan-generated synthetic data against (ii) models trained using standard supervised learning on 25 manually annotated experimental images, with **both models evaluated on the same independent experimental test set**.

This comparison reflects a realistic and commonly encountered scenario in AFM research, where manual annotation is costly and dataset sizes are inherently limited. The fact that SimuScan-trained models achieve performance comparable to—and in several cases exceeding—that of models trained on manually labeled experimental data provides strong evidence that the learned representations are not overfitted to synthetic artifacts but instead capture transferable, physically meaningful features of experimental AFM images.

While the number of manually labeled experimental images per morphology remains limited (25 images per class), the evaluation is performed at the **instance level**, yielding 756 geometric objects and 1,124 bacterial instances. This provides a statistically meaningful evaluation set for object-level and pixel-level metrics.

Moreover, the intent of this study is not to claim exhaustive generalization across all possible AFM modalities or experimental conditions. Rather, our goal is to demonstrate that **synthetic data generated by SimuScan can replace or substantially reduce the need for manual annotation**, while achieving performance on experimental data that is competitive with traditional, annotation-driven training approaches. The head-to-head comparison in Table 1 directly supports this claim.

To avoid overstatement, we have revised the manuscript to explicitly frame these results as demonstrating **practical and task-relevant synthetic-to-experimental transfer**, rather than universal generalization. We believe this positioning accurately reflects both the evidence presented and the intended contribution of the work.

• Performance metrics in Table 1 appear to be computed primarily on synthetic test sets.

All reported performance metrics in Table 1—including AP@0.5, IoU, precision, recall, and F1 score—are computed **exclusively against manually annotated experimental AFM images**, with no SimuScan-generated data included in the evaluation sets. We have revised the text to explicitly state this and avoid any ambiguity regarding the origin of test data. Where synthetic datasets are used, this is now clearly limited to training only.

• The evaluation on real data (DNA, bacteria) is qualitative and anecdotal.

Regarding the evaluation on real AFM data, we would like to clarify that table 1 includes **explicit quantitative benchmarking on experimental datasets**, rather than relying solely on qualitative demonstrations. Model performance is evaluated on independently annotated experimental AFM images of various bacterial cells and nanostructured shapes using standard metrics (AP@0.5, IoU, precision, recall, and F1-score). These metrics are computed by directly comparing model predictions against manually labeled experimental ground truth and provide objective evidence of synthetic-to-experimental transfer. Analysis is performed both at the feature and global level.

In the new manuscript we also include qualitative examples shown for DNA (linear vs circular), DNA origami, microfabricated ORNL logo, and bacteria on non-planar substrates. These are intended to complement—rather than replace—the quantitative results by illustrating representative outputs and failure modes in realistic experimental contexts. As you will see, the qualitative demonstrations involve highly diverse or irregular morphologies for which comprehensive manual annotation would be impractical or counter to the goal of minimizing human intervention. Instead, these experiments serve as **out-of-distribution stress tests**, demonstrating that models trained exclusively on SimuScan data continue to operate robustly across markedly different length scales, surface topographies, and structural classes. Importantly, these qualitative demonstrations are not used to claim optimized task-specific performance, but

rather to assess **robustness, failure modes, and practical usability** of the approach under realistic experimental conditions.

- The introduction and results sections are well written but somewhat repetitive and promotional, particularly around AI/AFM "revolution" claims.
- Statements like "redefining what is achievable" and "establishes a foundation for self-guided experimentation" could be toned down.
- The paper would benefit from clearer differentiation between demonstrated and aspirational capabilities.

We sincerely thank the reviewer for highlighting these points and curbing our enthusiasm for this approach. We have carefully revised the entire manuscript to reduce repetition and to tone down promotional language. Statements that could be interpreted as aspirational have been either removed or reframed to clearly distinguish between **capabilities demonstrated in this work** (e.g., semi-autonomous feature detection and targeted imaging) and **future directions** (e.g., fully autonomous closed-loop experimentation). We now explicitly frame SimuScan as a foundation that enables such developments, rather than claiming them outright.

> Personally, I would recommend including quantitative evaluation on independent experimental datasets, with metrics such as IoU, mAP@[0.5:0.95], and F1 scores across several morphologies.

We thank the reviewer for this recommendation. This has now been fully addressed in the revised manuscript. Specifically, we include quantitative benchmarking on independent experimental AFM datasets across multiple morphologies (bacterial cells, and engineered nanostructures), reporting standard metrics including IoU, mAP@[0.5:0.95], precision, recall, and F1-score (see Table 1 and Section 2.2). These evaluations are performed on experimental images not used during training or hyperparameter selection, providing objective evidence of synthetic-to-experimental transfer.

Taken together, these revisions substantially strengthen the evidentiary basis for the manuscript's central claims and clearly delineate demonstrated performance from forward-looking potential. We believe the revised manuscript now provides a balanced, quantitatively grounded assessment of SimuScan's capabilities and limitations.

We have included in the text:

Model performance was evaluated using an independently prepared ground-truth dataset consisting exclusively of experimental AFM images; no SimuScan-generated images were included in the evaluation set. The dataset comprised AFM images of geometric objects as well as two microbial morphologies—bacilli (rod-shaped) and cocci (spherical)—with manual annotations performed using Roboflow. Although the number of manually labeled experimental

images was limited (25 images per sample type), this was sufficient because individual AFM images typically contain a high density of repeated object instances. Standard data augmentation was applied to introduce controlled variability in orientation, scale, and contrast while preserving underlying morphology. As a result, the evaluation dataset contained 756 geometric object instances and 1,124 bacterial instances, providing a quantitatively informative and morphology-diverse benchmark for assessing segmentation performance.

We intentionally did not pursue extensive training or fine-tuning on large numbers of manually annotated experimental AFM images, as the objective of this study was not to maximize task-specific performance through annotation-intensive workflows. In prior AFM segmentation studies using YOLO- or U-Net–based architectures, performance is often improved by training on hundreds of manually labeled experimental images—a process that typically requires days to weeks of expert effort due to careful feature delineation, instance separation, and artifact rejection. In contrast, the central aim of this work is to evaluate whether SimuScan-generated synthetic data can substantially reduce or eliminate this manual labeling burden while still enabling effective model training. In support of this, qualitative inspection of predictions from the manually trained model further confirmed that the level of performance achieved was sufficient for the objectives of the present study, namely assessing the Simuscan data generalization rather than maximizing any task-specific accuracy.

While the authors describe SimuScan as unique, no comparisons are made with existing synthetic AFM data generators (e.g., BioAFMviewer, Degenhardt et al.). Without such benchmarks, it is unclear whether SimuScan provides superior realism or utility for training. Adding a benchmark against at least one existing synthetic dataset or model trained on small real datasets to quantify the improvement due to SimuScan.

We thank the reviewer for this suggestion and agree that it is important to clearly position SimuScan relative to prior work on synthetic AFM data. We have revised the manuscript to clarify this distinction more explicitly.

Existing tools such as **BioAFMviewer** and the framework by **Degenhardt et al.** are designed for fundamentally different objectives than SimuScan. BioAFMviewer generates molecular-scale AFM images from atomistic structural models to aid structural interpretation, while the HORNET framework targets RNA structure reconstruction under controlled noise assumptions. These methods operate in different resolution regimes, rely on structure-specific or atomistic priors, and are not intended to generate large, diverse, fully labeled datasets for training general-purpose deep learning models.

In contrast, SimuScan is designed as a **mesoscale, physics-informed synthetic data engine** for training segmentation and feature-identification models that must generalize across heterogeneous samples, substrates, and imaging conditions. Given these differences in modeling assumptions, scale, and intended use, we believe direct benchmarking against these tools would not be meaningful.

Instead, we evaluate SimuScan based on its intended utility: enabling models trained exclusively on synthetic data to generalize to experimentally acquired AFM images. Accordingly, validation focuses on quantitative and qualitative synthetic-to-experimental transfer across independent experimental datasets and diverse morphologies. We have revised the manuscript to clarify this scope and to articulate evaluation criteria aligned with SimuScan's intended application.

Table 1 – Informative yet difficult to interpret. Reformat as a clear matrix with error bars or replace colored dots with shaded highlights for print clarity. A complementary bar chart could visualize performance differences.
Figure 3 – The closed-loop workflow schematic is intuitive but lacks experimental imagery. Numbered process steps and inclusion of a real scan inset would make it more instructive.
Figure 4 – Impressive quantitative analysis of P. aeruginosa cells. Add scale bars, representative zoom-ins, and explicit sample size (n = 100) with labeled units on plots. Histogram fits should list model parameters.
Figure 5 – The large-area mapping of mixed bacterial populations is visually striking but needs validation of bacterial identity (e.g., optical overlay) and explicit scales for density maps and insets.
Across all figures, maintain uniform color scales, include scale bars, and use accessible color palettes for color-blind readers (just a suggestion)

We thank the reviewer for their careful and constructive feedback on the figures and tables. We greatly appreciate the level of detail provided, and we have undertaken a substantial revision of the visual materials throughout the manuscript in response. Many figures have been redesigned, split, or expanded for clarity; color palettes and scale bars have been standardized; and new experimental insets, annotations, and quantitative comparisons have been added where appropriate. Table 1 has also been reformatted to improve readability, including clearer layout, consistent units, and expanded descriptions.

Our goal throughout the revision process has been to incorporate the reviewer's suggestions as fully as possible and to enhance the clarity, accessibility, and scientific rigor of the visual presentation. We hope that these improvements meet the reviewer's expectations and contribute to a clearer and more effective communication of the results.

The "autonomous AFM" section presents a compelling integration of real-time segmentation and stage control, yet the process still involves human-defined parameters, thresholds, and stopping criteria. The absence of discussion of failure modes (e.g., misdetections, tip crashes, or drift correction) overstates the level of autonomy. The SPM community wants to know! By recasting this section as a semi-autonomous or guided AFM workflow, and explicitly describing system limitations and failure recovery strategies.

We thank the reviewer for this insight which we feel is an important observation. We fully agree that, although our long-term objective is to enable fully autonomous AFM, the present implementation still requires user-defined parameters, thresholds, and stopping criteria, and therefore should be characterized as a **semi-autonomous** or **guided** workflow rather than a fully autonomous system. The reviewer is also correct that explicitly discussing limitations and potential failure modes is essential for transparency and for informing the SPM community about realistic operational performance.

In response, we have revised this section of the manuscript to clearly frame the approach as a **semi-autonomous closed-loop workflow**, and we now provide a dedicated discussion of system limitations, including potential misdetections, tip-related failures, drift correction considerations, and strategies for recovery or user intervention. We believe this reframing more accurately reflects the current state of the system while still highlighting the significant hurdle overcome—namely, that models trained entirely on synthetic data can successfully guide real-time AFM operations across diverse sample types.

We thank the reviewer again for encouraging this clarification, which we believe improves both accuracy and relevance for the AFM community.

Although SimuScan includes key AFM-specific artifacts (tip convolution, noise, flattening), their quantitative accuracy is not evaluated. It is unclear whether the simulated distortions statistically resemble real AFM data. This point is also important to know for the end-users. Providing a quantitative comparison between simulated and experimental AFM images, for instance, using height–height correlation functions, noise spectra, or power spectral density analysis will certainly be helpful.

In the present work, however, our goal is not to reproduce a statistically exact digital twin of a specific AFM instrument, but rather to generate **artifact-informed synthetic data with sufficient realism and variability to enable effective training and generalization of deep learning models**.

Accordingly, SimuScan models AFM artifacts at the **mechanistic and phenomenological level**—including tip–sample convolution, electronic and 1/f noise, line-related distortions, and flattening effects—using parameterized representations that span ranges commonly observed in experimental practice. The primary validation criterion we adopt is therefore **synthetic-to-experimental transfer performance**: models trained exclusively on SimuScan data are evaluated on independent experimental AFM images, where successful generalization provides an integrated, task-relevant measure of whether the simulated artifacts capture the dominant statistical and structural features that matter for downstream analysis.

We agree that detailed statistical matching of individual artifact distributions (e.g., PSDs or correlation functions) would be informative for instrument-specific digital-twin applications. However, such analyses would require conditioning SimuScan on a particular AFM system, probe, and operating mode, which is beyond the scope of this study and would reduce generality. Instead,

SimuScan is intentionally designed to expose models to **broad, controllable ranges of artifact realizations**, improving robustness across instruments and experimental conditions rather than optimizing fidelity to a single setup.

To address the reviewer's concern and improve clarity for end users, we have revised the manuscript to explicitly state this design philosophy and to clarify that SimuScan prioritizes **training utility and generalization robustness** over strict statistical replication of a specific instrument. We also note in the revised discussion that incorporating quantitative, instrument-specific artifact calibration (e.g., PSD- or correlation-based matching) represents a natural extension of the framework for future studies focused on digital-twin fidelity.

The paper states that SimuScan is "available upon request," which contradicts the claim of open science. Reproducibility depends on full code and configuration availability. My recommendations will be to publicly release the full SimuScan source code and documentation (or a Docker container) to ensure replicability, even though some parts are on GitHub and I haven't tested it so far … but I certainly will!

We thank the reviewer for raising this important point regarding reproducibility and open science. We fully agree that transparency, accessibility, and reproducibility are essential for advancing machine-learning–enabled AFM research, and it is our intention to make SimuScan broadly available to the community.

At present, portions of the SimuScan framework are undergoing institutional intellectual property review, with a patent application in preparation and discussions with an AFM vendor are ongoing. Until this process is complete, we are not permitted to publicly release the full source code. For this reason, the software is currently described as "available upon request," which allows us to share the framework with researchers under appropriate licensing during the IP review period.

To support reproducibility during this interim phase, we have:

- publicly released all **non-restricted components** of the codebase on GitHub,
- provided **detailed configuration files, example datasets, and documentation** describing simulation parameters, model training, and evaluation workflows, and
- ensured that **all experiments reported in the manuscript can be reproduced** using the released components and documented configurations.

We have revised the manuscript to clarify this status and to avoid ambiguity regarding availability. Upon completion of the IP review, we plan to release the **full SimuScan package**, including comprehensive documentation and containerized environments, to enable straightforward replication and adoption by the community. In the meantime, we would be pleased to provide access to interested researchers under a research license, and we welcome the reviewer's interest in testing the framework.

The bacterial and DNA demonstrations are visually appealing but lack quantitative biological validation. The models' specificity (ability to distinguish true cells from debris) and robustness to sample variation remain unclear.
Maybe the authors could include, in SI section, confusion matrices or precision/recall data for biological classifications, ideally compared to manual expert annotation.

In the present study, our primary objective is not biological classification per se, nor the maximization of task-specific performance for any single sample type, but rather to evaluate whether models trained exclusively on synthetic data can robustly detect and segment nanoscale features across diverse experimental contexts. Accordingly, we focus our quantitative evaluation on controlled experimental benchmarks (Table 1), where manual annotation is well defined and enables reliable computation of precision, recall, F1-score, and IoU on real AFM images. These metrics provide the principal quantitative evidence supporting synthetic-to-experimental generalization.

For additional demonstrations involving DNA, bacterial samples on pillar substrate and nanostructured samples, we therefore emphasize **qualitative robustness and specificity** under realistic imaging conditions, including heterogeneous substrates, surface debris, and morphological variability. In these experiments, model predictions were consistently assessed against expert interpretation, and representative examples—illustrating both correct discrimination between biological features and background structures as well as typical failure modes—are presented. We now clarify this distinction explicitly in the revised manuscript to avoid the impression that these demonstrations are intended as fully quantitative biological classification benchmarks.

Importantly, this design choice reflects the broader goal of SimuScan: to enable **rapid, annotation-light deployment** of AFM automation workflows, rather than annotation-intensive optimization that can require days to weeks of expert effort and still remain subject to human bias and inter-annotator variability. We believe that this combination of rigorous quantitative benchmarking where ground truth is well defined, together with qualitative stress testing on biologically complex datasets, provides an appropriate and transparent assessment of model performance for the scope and aims of this study.

Minor Comments

1. Terminology: In Figure 4, "Bacillus cells" likely refers to Pseudomonas aeruginosa; please correct to avoid taxonomic confusion.
2. Metrics: Use consistent metrics (preferably mAP@[0.5:0.95]) across models for fair comparison.
3. Figures: Figures 2–5 contain excessive subpanels and dense captions; consider streamlining for readability.
4. Data Volume: Clarify how 5,000 synthetic images suffice for training deep architectures—were

data augmentations or tiling used?
5. Manuscript Tone: Several statements (e.g., "redefining what is achievable") are overly promotional; suggest tempering to reflect demonstrated results.
6. Formatting: Minor typographical inconsistencies (μ vs μ, superscripts, spacing) should be corrected.

We thank the reviewer for these helpful minor suggestions and have addressed each point accordingly. Terminology, formatting, and tone have been corrected; figures and captions streamlined; metrics standardized to mAP@[0.5:0.95]; and additional clarification added regarding training data volume and augmentation. We appreciate the reviewer's careful attention to detail, which has improved the clarity and consistency of the manuscript.

In conclusion, this work presents an exciting and potentially field-shifting approach to autonomous AFM. However, its claims of synthetic-to-real generalization and full autonomy require stronger experimental evidence, quantitative benchmarking, and improved reproducibility. With these revisions, the paper could make a substantial contribution to autonomous nanometrology and AI-driven instrumentation.

I recommend major revision but I strongly believe that by addressing my comments and remarks, the paper will be among the key-ones in this growing field merging SPM and AI.

We thank the reviewers for their thoughtful summary and for recognizing the potential impact of this work on autonomous AFM and AI-driven instrumentation. We especially appreciate the reviewers call for stronger experimental validation, quantitative benchmarking, and enhanced reproducibility. In response to the reviewer's suggestions, we have undertaken substantial revisions; including expanded experimental datasets, new quantitative evaluations, methodological extensions and we are confident that that paper is now ready for publications.
We are grateful for the reviewer's constructive guidance and share their optimism that the revised manuscript will make a meaningful contribution to the emerging intersection of SPM and AI.

Reviewer #3 (Remarks to the Author):

Millan-Solsona et al present a novel tool, SimuScan, to create artificial AFM datasets based on specific criteria. They show that autonomous operation of AFM based on artificially-trained machine learning (ML) is possible over a range of samples, from simple geometrical patterns on hard substrates to bacteria. They also emphasise the analysis capabilities of ML tools trained on with SimuScan, as illustrated on DNA structures.

The authors argue that the absence of tractable and consistent datasets for ML is the largest bottleneck to wide adoption of ML strategies in AFM. I agree with the authors about this being a major roadblock, both for automation and image interpretation. Several attempts to create extensive AFM databases for ML have not succeeded for lack of suitable large sets and/or metadata information. Most successful publications tend to focus on a narrow and well defined problem.

In this context, SimuScan clearly offers a step in the right direction. However, it remains to be seen whether this step is incremental or a step-change.

Response:
We sincerely thank the reviewer for the detailed reading and the positive comments on our contribution. We are pleased that the reviewer highlighted both the value of SimuScan and the relevance of the challenge we aim to address—namely, the lack of large, consistent, and well-annotated datasets for training machine-learning models in AFM. We fully agree that this limitation has hindered the development of robust automation and ML-based analysis strategies within the AFM community for many years.

We also appreciate the reviewer's remark that SimuScan represents a step in the right direction. The primary goal of our work is precisely to provide a flexible, physically informed, and easily extensible tool capable of overcoming current data-availability constraints and enabling new AI-based workflows. In the revised manuscript, we have expanded the discussion to better reflect this context and to clarify in which ways SimuScan constitutes a step toward more advanced automation and analysis capabilities. We have also toned down the language to clearly separate capabilities demonstrated in this work from future directions.

In response to the reviewer's question regarding whether this represents an incremental improvement or a qualitative change. We have added additional components to the SimuScan package, including a experimental self seeded approach which we are excited about and significantly expanded the demonstrations by including additional experiments. We have additional text in the manuscript explicitly outlining the current limitations of our approach, as well as its potential to evolve toward more complex scenarios as additional simulation modules are incorporated (e.g., more advanced tip–sample interactions, digital twins, spectroscopic modes, nonlinear dynamics). We hope that this clarification helps readers accurately position the present scope and future trajectory of the tool.

My main concern with SimuScan as presented here is that it is illustrated on clearly defined and easily identified criteria for samples such as geometrical shapes or loop vs elongated strand. While

likely sufficient for large (>100 nm) samples, I am not convinced that it is enough for nanoscale images where countless known and unknown effects may be at play, from convolution to intermittent molecular contamination of the tip, asymmetric tip shapes, unstable sample fixation, dirty sample, inappropriate scanning parameters, etc. I feel this is important because the power of AFM lies at the nanoscale, where optical approaches are less straightforward.

We thank the reviewer for raising this important and technically substantive point. We agree that nanoscale AFM image formation is governed by many interacting—and often poorly characterized—factors, including asymmetric tip convolution, intermittent contamination, unstable sample fixation, and suboptimal scan parameters. We also agree that in the previous version of the manuscript our examples could be seen as those of simple samples such as geometrical shapes or loop vs elongated strand which could be represented as analytically defined shape models which is not universally the case.

To directly address this challenge, we have significantly expanded SimuScan in the revised manuscript to support **non-parameterized, morphology-driven synthetic data generation**. Specifically, SimuScan now includes self-seeded geometry workflows in which either an experimentally acquired AFM images or externally defined CAD/STL geometries are used directly as the morphological seed for the subsequent synthetic dataset generation. In this mode, more complex shapes, including sometimes unknown nanoscale effects present in real AFM data are implicitly embedded in the training data without requiring explicit modeling of each contributing mechanism. This capability substantially extends SimuScan beyond the simple geometric and topology-based examples shown in the initial submission.

The authors show that SimuScan can distinguish different DNA shapes, but this is still a relatively simple question regardless of the possible issues mentioned above. What about distinguishing the orientation of an adsorbed (small) molecule? Or a clean from a dirty tip? Generally, how can realistic nanoscale data be generated without knowing all the details about the sample and the tip-sample interactions that can dramatically change the resulting image? Is this a limitation of SimuScan?

Response:

We sincerely thank the reviewer for their insights. The initial demonstrations using simple geometric features and DNA topology were intentionally chosen as controlled benchmarks with well-defined ground truth. In response to the reviewers valid concerns, we have expanded the manuscript to include substantially more complex experimental systems, including bacterial cells imaged on nanopatterned pillar substrates and DNA origami nanoassemblies, which introduce non-planar backgrounds, increased morphological complexity, and realistic substrate effects.

We further acknowledge that more dynamic or intermittent phenomena; such as time-dependent molecular contamination, unstable adsorption, or poorly optimized scan parameters, are inherently difficult to capture in their entirety. Rather than attempting exhaustive modeling upfront, **SimuScan is intended as a robust and rapid starting point**: in cases involving more complex interactions or requiring very high levels of task-specific precision, we fully expect that tuning of

the SimuScan data generator, as well as model architecture and hyperparameters, will be necessary, and the framework has been explicitly designed with this flexibility in mind. However, such task-specific optimization is not the goal of the present work. Instead, SimuScan is designed as a modular and extensible framework, allowing additional effects to be incorporated progressively as needed.

Importantly, we want to stress that SimuScan's primary objective is not to reproduce every nanoscale interaction with quantitative fidelity, instead we aim to enable robust, annotation-free **model training under uncertainty** by exposing deep learning models to dominant AFM artifact mechanisms **sampled across broad, physically plausible ranges**.

We therefore view the reviewer's concern as clarifying the intended scope of SimuScan rather than identifying a fundamental limitation. SimuScan is not designed to solve all nanoscale AFM interpretation problems, but to provide a flexible, generalizable pathway toward automated feature detection across diverse, imperfect, and evolving experimental conditions. Consistent with this perspective, we have also tempered the manuscript's tone to present SimuScan as a significant step toward scalable synthetic datasets and semi-autonomous AFM workflows, rather than a complete or final solution.

A second concern/question is the apparent lack of spectroscopic information. Currently, most AFM operate in spectroscopy mode (peak force tapping, wave mode, etc) with considerable information in the spectroscopy curve obtained at each pixel. Can SimuScan create relevant information to train ML of the spectroscopy curves?

We thank the reviewer for raising this important question. SimuScan, in its present form, is focused exclusively on generating realistic **morphological (topographic) datasets** and does not simulate full spectroscopy curves such as those obtained in peak-force tapping, WaveMode, or force–distance spectroscopy. While our approach is compatible with topography channels derived from these techniques, the framework does not currently aim to reproduce the underlying per-pixel spectroscopic signals.

We have clarified this scope in the revised manuscript.

While SimuScan captures a broad range of AFM-specific artifacts relevant for realistic synthetic data generation, this first implementation makes deliberate scope choices leaving clear avenues for future extension. At present, the framework focuses on the topographic channel, which is broadly applicable across common AFM modes (e.g., dynamic, static, PeakForce Tapping, and wave-based operation), while additional contrast channels such as phase, amplitude, or spectroscopic signals are not explicitly modeled. Long-timescale thermal and mechanical drift is also not simulated, reflecting the prevalence of closed-loop AFM systems and standard post-processing procedures that mitigate these effects in many experiments. In addition, SimuScan

models static double-tip artifacts but does not yet capture intermittent or time-dependent probe contamination. Extending the framework to incorporate additional signal channels, drift dynamics, and evolving probe conditions represents a natural and modular path forward, and would further expand the realism, versatility, and applicability of SimuScan across a wider range of AFM modalities and experimental conditions.

At the same time, we agree that ML applied to spectroscopy curves represents an exciting direction, and SimuScan-trained segmentation models already enable precise identification of features or regions of interest where spectroscopy could be targeted. In future extensions, we plan to incorporate modules for generating synthetic functional property maps and, ultimately, spectroscopic curves, building on the same modular architecture.

The presentation of the paper is generally clear, but Fig 2 contains too much information (and too small in h/i) to be easily appreciated.

We thank the reviewer for this helpful suggestion. We agree that the original Fig. 2 contained too much information and that panels h/i were difficult to appreciate at the provided scale. In the revised manuscript, we have reorganized the figures by separating the pipeline schematic (new Figure 2) from the dataset examples (new Figure 3), enlarging key panels, and simplifying the layout to improve readability and visual clarity.