ADAPTING GENERAL-PURPOSE FOUNDATION MODELS FOR X-RAY PTYCHOGRAPHY IN LOW-DATA REGIMES

AI for Accelerated Materials Design: 39th Conference on Neural Information Processing Systems (NeurIPS 2025)

Robinson Umeike[1,*], Neil Getty[2], Xiangyu Yin[2] and Yi Jiang[2]
[1]The University of Alabama, [2]Argonne National Laboratory
[*] Work performed while at Argonne National Laboratory

ABSTRACT SUMMARY

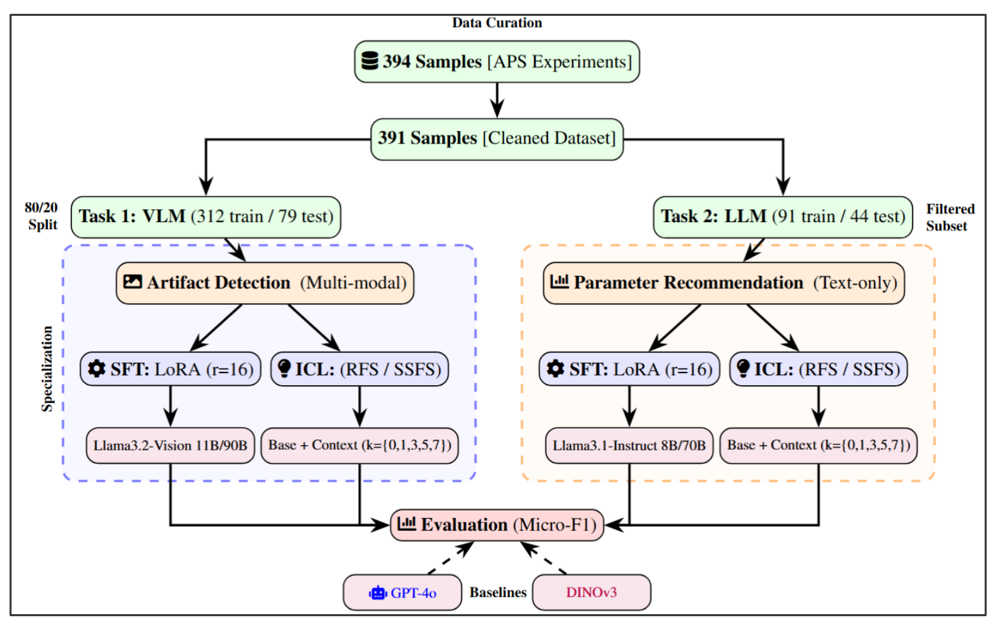
Adapting foundation models for specialized scientific tasks is critical, yet the optimal strategy remains unclear. We introduce **PtychoBench**, a multi-modal benchmark for ptychography, to systematically compare Supervised Fine-Tuning (SFT) versus In-Context Learning (ICL) in a data-scarce regime. Our findings reveal the optimal pathway is task-dependent. For visual tasks (VLMs), SFT and ICL are **complementary**, achieving the highest **mean performance (Micro-F1 0.728)**. Conversely, for textual tasks (LLMs), ICL on a base model is **superior (mean Micro-F1 0.847)**, outperforming a "super-expert" SFT model **(mean Micro-F1 0.839).** Benchmarked against **GPT-4o** and **DINOv3**, our results on this dataset highlight that visual tasks benefit from hybrid specialization, while textual reasoning favors flexible base models in science-based agentic systems.

CONTRIBUTION

- ❑ **Novel Benchmark:** 391 expert-annotated samples from the Advanced Photon Source (APS)
- ❑ **Dual Tasks:**
  - ✓ **Task 1 (Visual):** Artifact Detection (VQA).
  - ✓ **Task 2 (Textual):** Parameter Recommendation (QA).
- ❑ **Data-Scarce Regime:** Limited training data available, mimicking real experimental constraints.

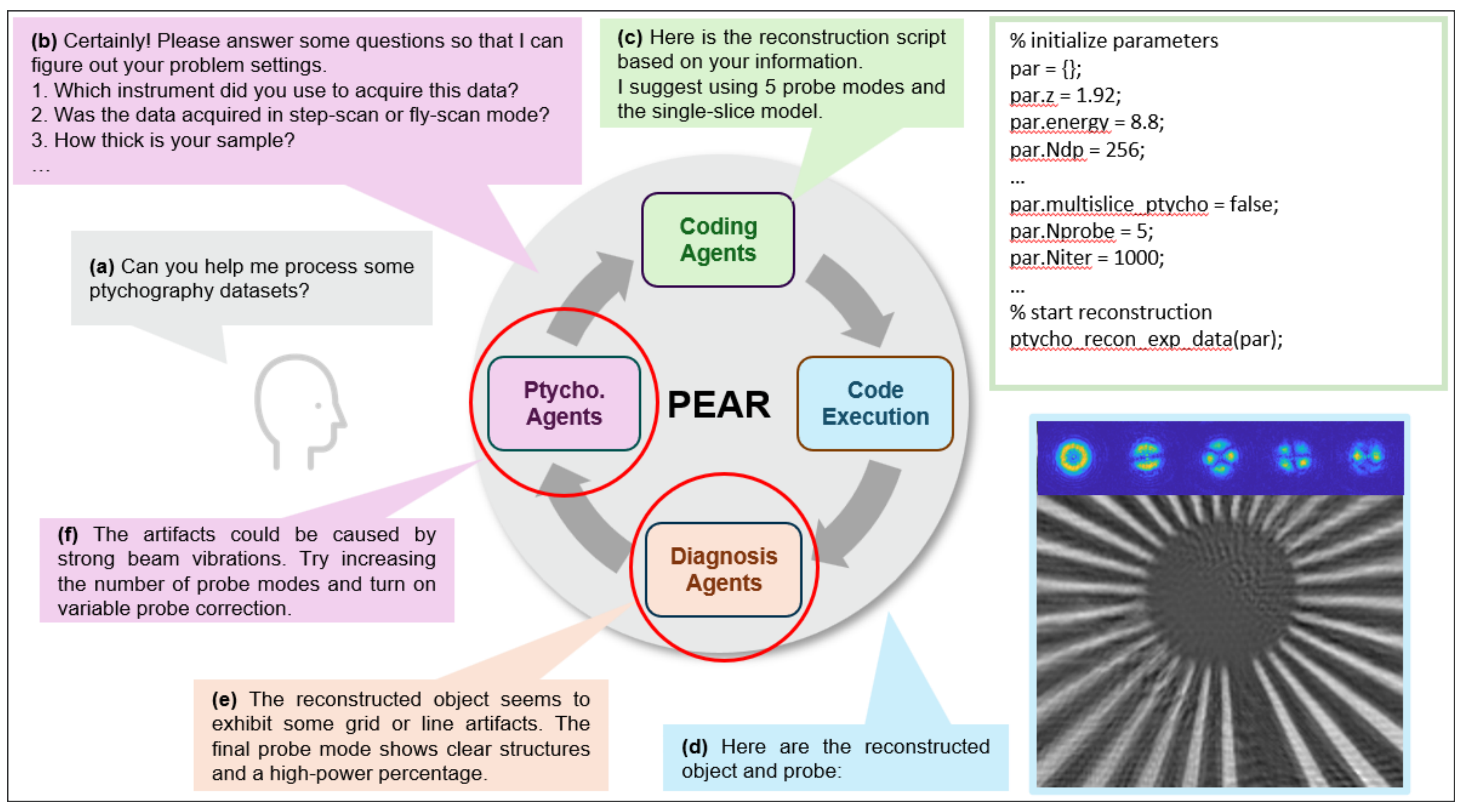
METHODS

- ❑ **Models:** Llama 3.2 (11B/90B *Vision*), Llama 3.1 (8B/70B *Text*).
- ❑ **ICL Strategy:** Compared Random Few-Shot (RFS) vs. Sample-Specific Few-Shot (SSFS).


Data Curation
394 Samples [APS Experiments]
391 Samples [Cleaned Dataset]
80/20 Split
Task 1: VLM (312 train / 79 test)
Task 2: LLM (91 train / 44 test)
Filtered Subset
Specialization
Artifact Detection (Multi-modal)
Parameter Recommendation (Text-only)
SFT: LoRA (r=16)
ICL: (RFS / SSFS)
SFT: LoRA (r=16)
ICL: (RFS / SSFS)
Llama3.2-Vision 11B/90B
Base + Context (k={0,1,3,5,7})
Llama3.1-Instruct 8B/70B
Base + Context (k={0,1,3,5,7})
Evaluation (Micro-F1)
GPT-4o
Baselines
DINOv3

BACKGROUND & MOTIVATION

- ❑ **Context:** X-ray ptychography is a computational imaging technique that achieves high resolution but requires manual parameter tuning.
- ❑ **Problem:** General-purpose models struggle with this specialized scientific data, and *no standard benchmark exists* to evaluate them.
- ❑ **Goal:** Automate manual analysis using "Agentic Workflows" (e.g., PEAR), with VLM (Diagnosis) and LLM (Recommendation) agents.
- ❑ **Approach:** We systematically compare two adaptation paths: **Supervised Fine-Tuning (SFT)** vs. **In-Context Learning (ICL)**.



(a) Can you help me process some ptychography datasets?

(b) Certainly! Please answer some questions so that I can figure out your problem settings.
1. Which instrument did you use to acquire this data?
2. Was the data acquired in step-scan or fly-scan mode?
3. How thick is your sample?
…

(c) Here is the reconstruction script based on your information.
I suggest using 5 probe modes and the single-slice model.

```
% initialize parameters
par = {};
par.z = 1.92;
par.energy = 8.8;
par.Ndp = 256;
...
par.multislice_ptycho = false;
par.Nprobe = 5;
par.Niter = 1000;
...
% start reconstruction
ptycho_recon_exp_data(par);
```

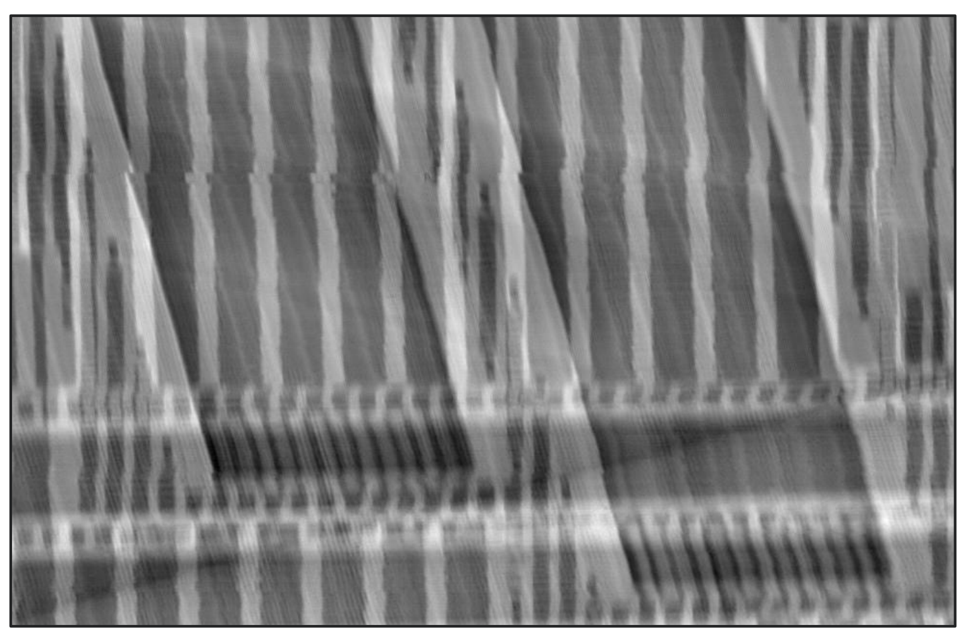
(d) Here are the reconstructed object and probe:

(e) The reconstructed object seems to exhibit some grid or line artifacts. The final probe mode shows clear structures and a high-power percentage.

(f) The artifacts could be caused by strong beam vibrations. Try increasing the number of probe modes and turn on variable probe correction.

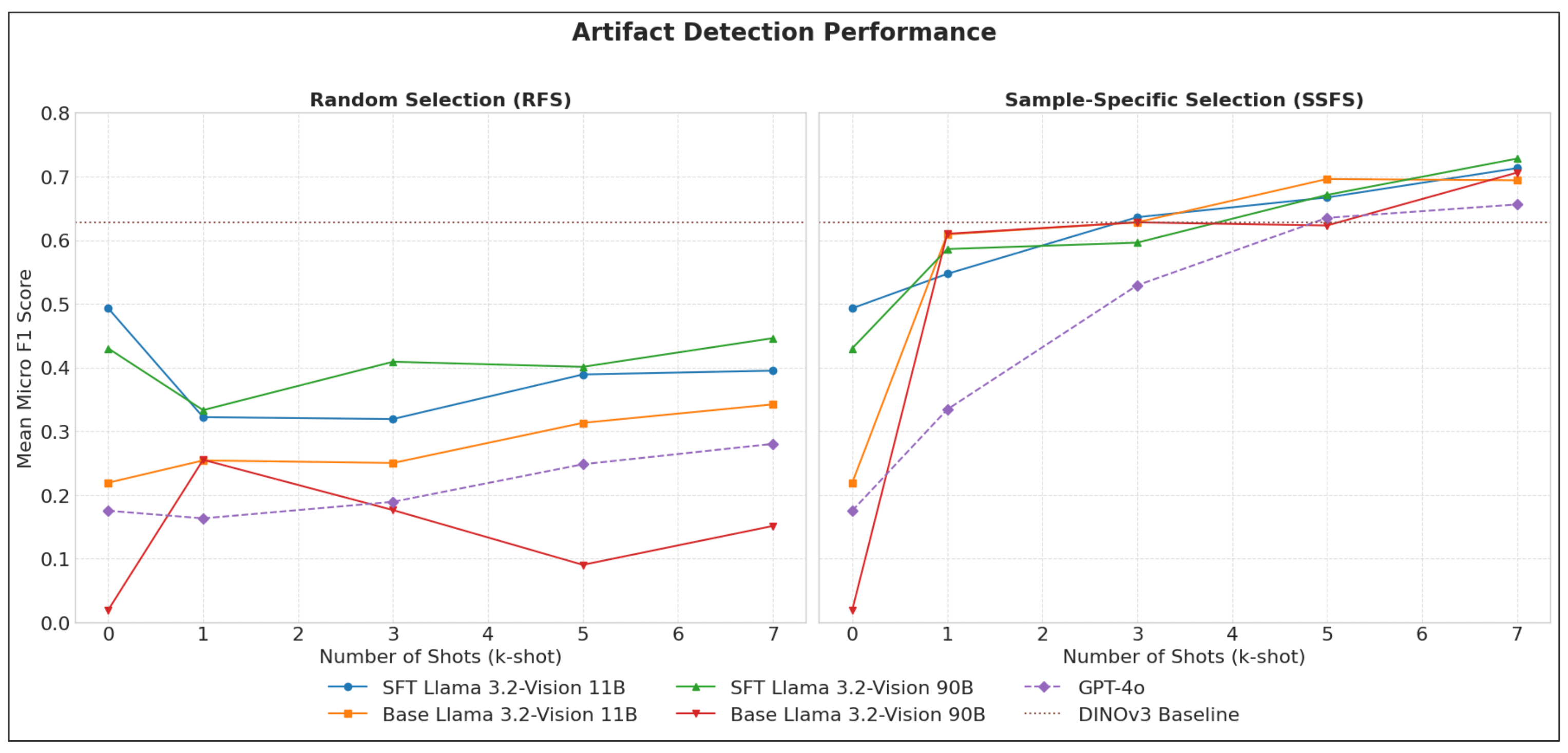
Sample Info:
- ✓ **Type**: Integrated circuit
- ✓ **Artifact**: Local Distortion

Expert recommendation:
- ✓ Increase the number of iteration.
- ✓ Reduce the size of diffraction patterns by a factor of 2
- ✓ Turn off variable probe correction.

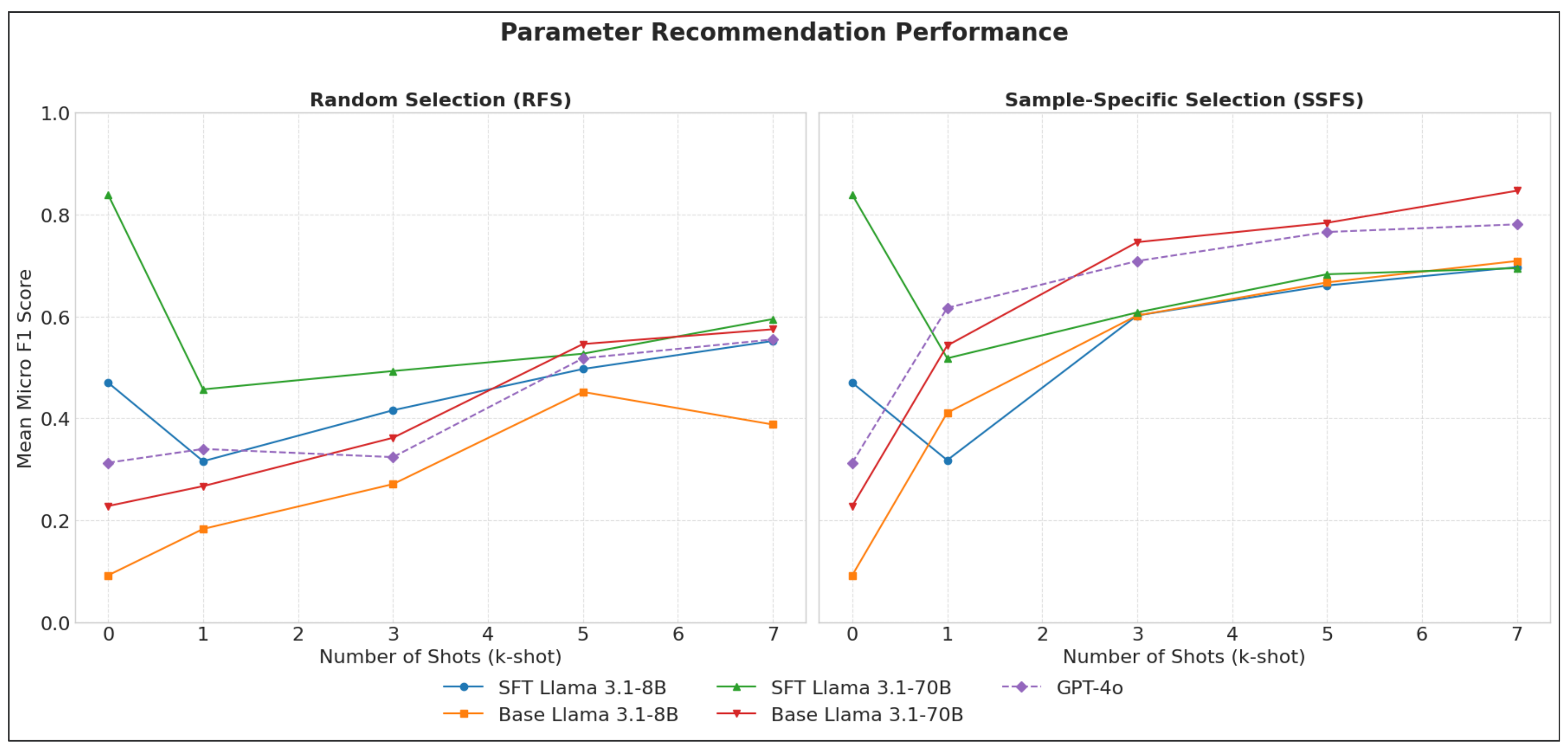


RESULTS

❑ **Visual diagnosis:** The winning strategy **combines SFT and SSFS (0.728 F1)**. Fine-tuning builds a "*visual vocabulary*" that context refines, allowing the larger **90B model** to overcome its lower standalone baseline and outperform the smaller expert when unlocked by relevant examples.



❑ **Textual recommendation:** In contrast, textual recommendation favors **a Base 70B Model with SSFS (0.847 F1)**. While SFT created a "*brittle super-expert*" that degraded with context, the flexible base model effectively utilized examples to surpass the specialized expert.



DISCUSSION & CONCLUSION

- ❑ **Contextual Interference:** We observed a consistent *contextual interference* phenomenon across fine-tuned models. While relevant context boosts scores, **random examples actively confuse the expert models**, causing performance to plummet below the zero-shot baseline (e.g., a 35% drop for SFT-11B), highlighting that retrieval quality is a critical safety constraint.
- ❑ **Data Efficiency:** Our results also highlight remarkable data efficiency. The Base-90B model, using **only 7 inference-time examples**, achieved a **mean F1 score (0.706)** statistically comparable to a **DINOv3** vision classifier trained on the entire dataset **(0.628).** This proves that guided foundation models can match traditional supervised learning with a fraction of the data.
- ❑ **Conclusion:** Our dataset reveals no one-size-fits-all strategy. Visual tasks (**artifact detection**) benefit from SFT priming, whereas textual tasks (**parameter recommendation**) require the flexibility of base models.

REFERENCES

- ❑ X. Yin, C. Shi, Y. Han, and Y. Jiang, (2024) *PEAR: A Robust and Flexible Automation Framework for Ptychography Enabled by Multiple Large Language Model Agents*.
- ❑ R. Umeike, N. Getty, F. Xia, and R. Stevens, (2025) *Scaling Large Vision-Language Models for Enhanced Multimodal Comprehension in Biomedical Image Analysis*. In 2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI), pp. 1--4.

Acknowledgement: This work was supported by the Laboratory Directed Research and Development (LDRD) Program at Argonne National Laboratory under Project Number 2025-0495. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357.

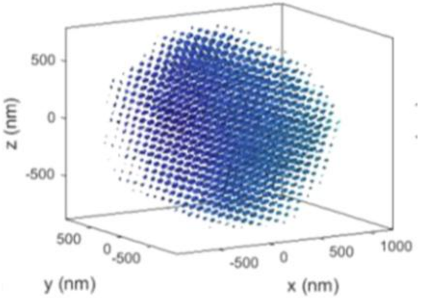
Have Questions? Please contact:
- ✓ ngetty@anl.gov (Code)
- ✓ yjiang@anl.gov (Dataset)
- ✓ *Scan for Code & Data Access Guidelines* ➔

SCAN ME

U.S. DEPARTMENT of ENERGY
Argonne National Laboratory is a U.S. Department of Energy laboratory managed by UChicago Argonne, LLC.

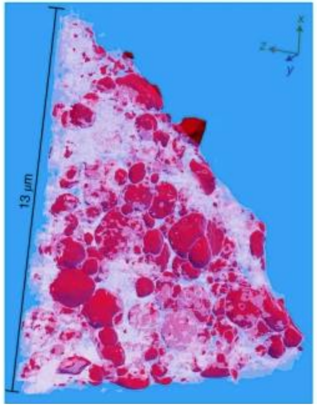
Transmission X-ray Ptychography @ APS

supercrystals



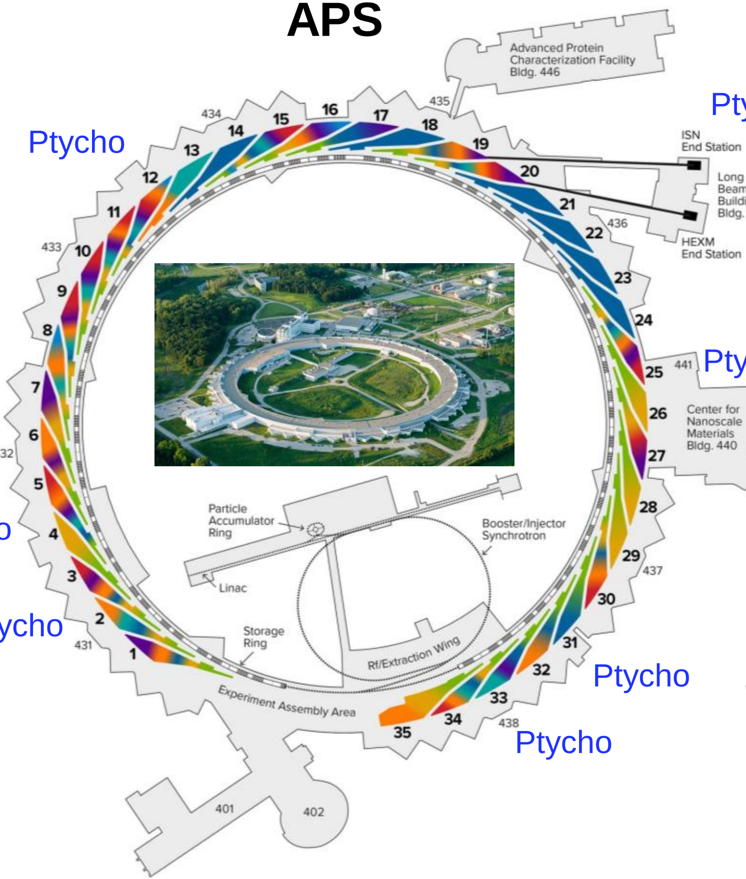
H. Calcaterra et al. *ACS Nano.* (2023)

liquid-metal particles



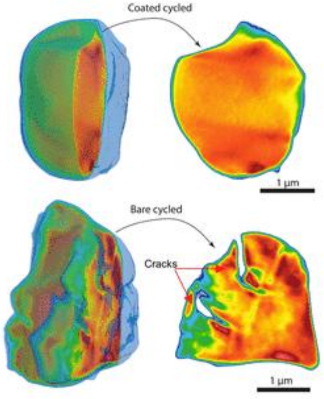
Y. Lin et al. *Nat. Chem.* (2022)

APS

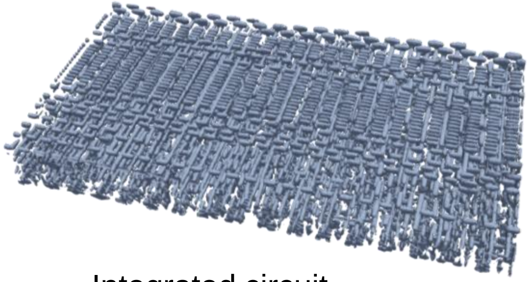


~10 instruments

Ni-Rich Cathodes



Q. Liu et al. *ACS Nano.* (2022)



Integrated circuit