Deep Learning-Based Segmentation of AFM Data for Epitaxial Graphene Analysis and SiC Terrace Identification

Khyati Sethia[1], Petr Strakoš[1], Martin Rejhon[2], Martin Papula[3], Ivana Ročňáková[3], Tomáš Musálek[3]

[1] IT4Innovations National Supercomputing Center, VSB Technical University of Ostrava
[2] Charles University
[3] Activair s.r.o.

INTRODUCTION

Atomic Force Microscopy (AFM) provides nanoscale information about surface structure and material properties. In epitaxial graphene grown on silicon carbide (SiC), surface regions differ based on:

- underlying SiC terraces (S2, S3)
- number of graphene layers (e.g., single-layer, bilayer)

Manual analysis of AFM images is: time-consuming, subjective, difficult to scale.

Goal: automate segmentation of AFM images into physically meaningful regions.

DATASET

Each AFM image contains information:

- Topography → surface height → identifies terraces
- Friction → mechanical response → indicates graphene layers
- Ground Truth Preparation: Converts colors to cluster IDs and then convert to boundary map.

Challenge:

- Overlapping signals between regions
- Limited annotated real data
- Difficult distinction (e.g., bilayer graphene on S2 vs S3)

DATASET PROPERTIES

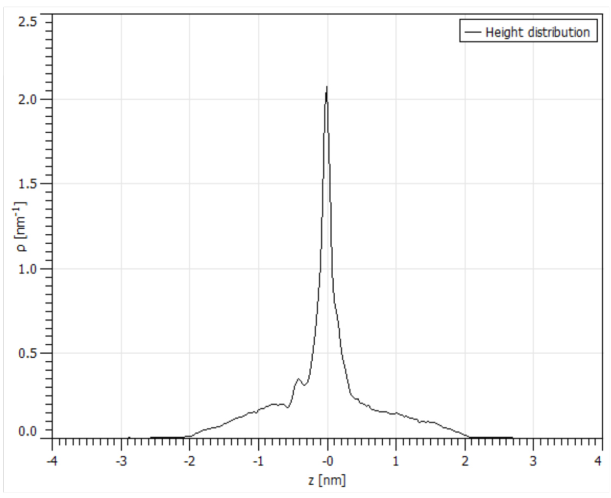

Figure: Height distribution

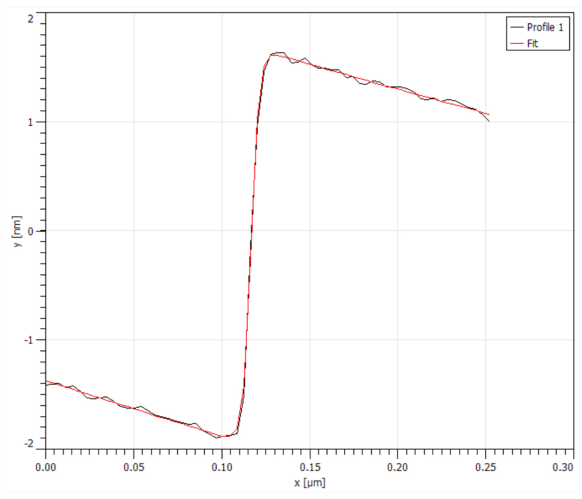

Figure: Fitting function to detect height differences between terraces.

Figure: Termination of SiC surface before epitaxial graphene growth (left) and after epitaxial graphene growth (right).

WORKFLOW

We propose a pipeline combining synthetic data generation + machine learning. Pipeline:

- Generate synthetic AFM data (Blender)
- Train segmentation model
- Apply model to real AFM images
- Compare and evaluate performancebu

SYNTHETIC DATA

To overcome lack of labeled data:

Surfaces are simulated using Blender

Controlled variations: terrace structures (step-like geometry), graphene layer distributions, noise and imaging artifacts.

Advantages:

- fully labeled data
- scalable dataset generation
- controlled experiments

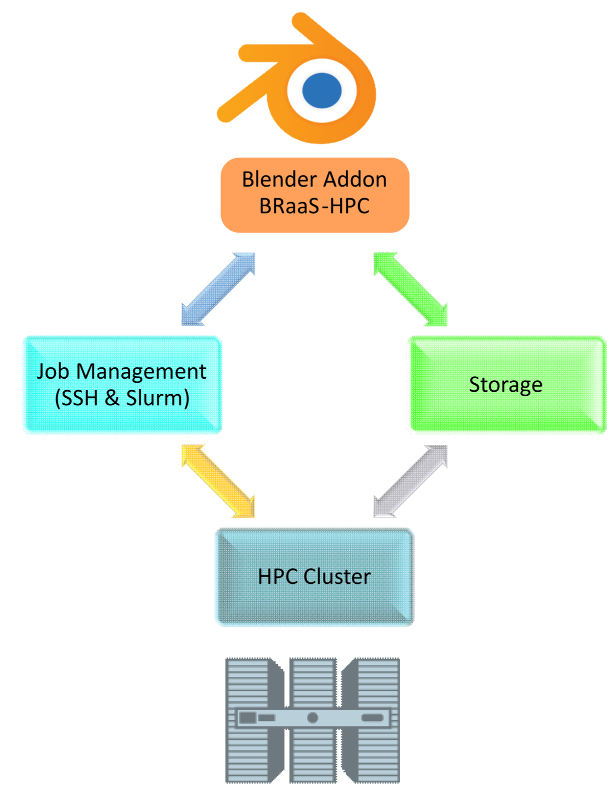

Figure: Distributed rendering using Blender addon and HPC

OVERVIEW OF THE PIPELINE

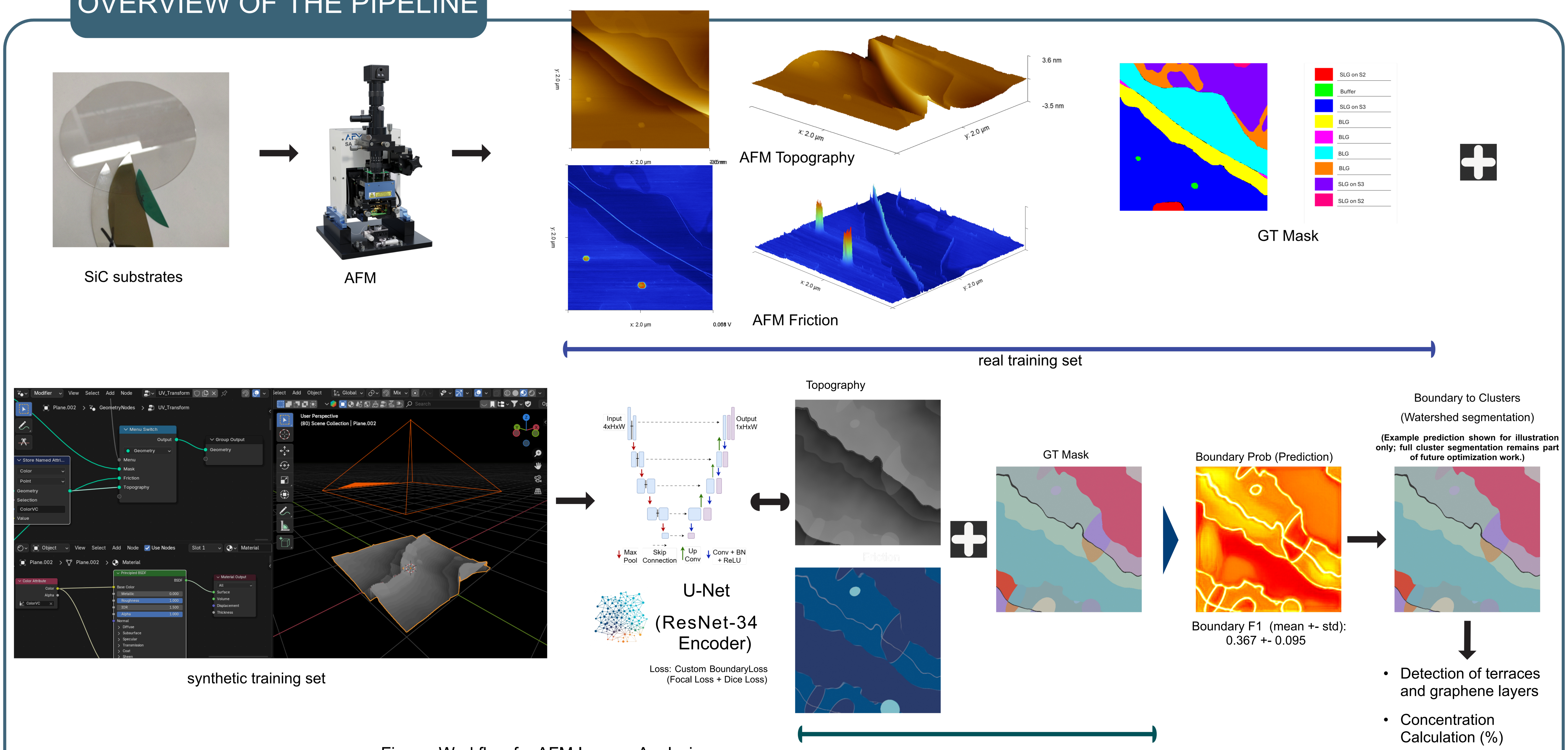

Figure: Workflow for AFM Images Analysis.

synthetic test set

CONCLUSION

- Boundary-aware learning framework for variable cluster counts. Uses those boundaries to reconstruct full clusters using watershed segmentation.
- Full region segmentation still requires further optimization (increased model capacity, expanded dataset, mltimodal learning, improved architectures. Instance-level cluster separation will be explored).
- Impact: This work establishes a foundation for AI-driven nanomaterial analysis, enabling high-throughput screening.

REFERENCES

- Blender Software: https://www.blender.org/, https://docs.blender.org/manual/en/latest/index.html
- Kocur, V., Hegrová, V., Patočka, M., Neuman, J., & Herout, A. (2023). Correction of AFM data artifacts using a convolutional neural network trained with synthetically generated data. Ultramicroscopy, 246. https://doi.org/10.1016/j.ultramic.2022.113666

ACKNOWLEDGEMENT

Project (FY01010025) - Research and development of advanced analysis and interpretation of atomic force microscopy (AFM) data using artificial intelligence and machine learning.