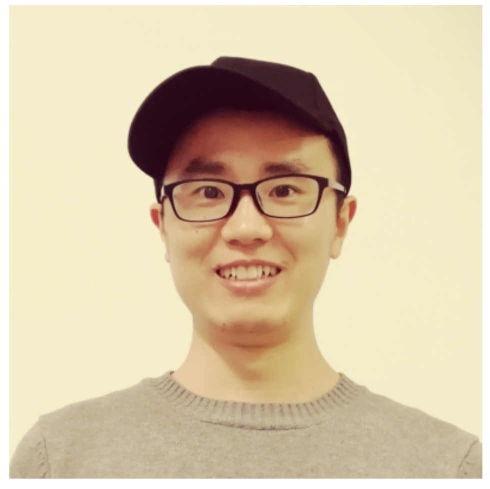
Surfaces and Interfaces at the Nanoscale (SIN)

Aalto University School of Science

Enhancing AFM Image Analysis Through Machine Learning with Style Translation and Data Augmentation



Jie Huang[a], Niko Oinonen[a,b], Lauri Kurki[a], Adam S. Foster[a,c]

[a]Department of Applied Physics, Aalto University, Finland
[b]Nanolayers Research Computing Ltd., London N12 0HL, United Kingdom
[c]WPI Nano Life Science Institute (WPI-NanoLSI), Kanazawa University, Japan

jie.huang@aalto.fi

Introduction

Atomic Force Microscopy (AFM) plays a crucial role in characterizing nanostructures. The **Probe Particle Model (PPM)** [1] offers an efficient method for simulating AFM images.

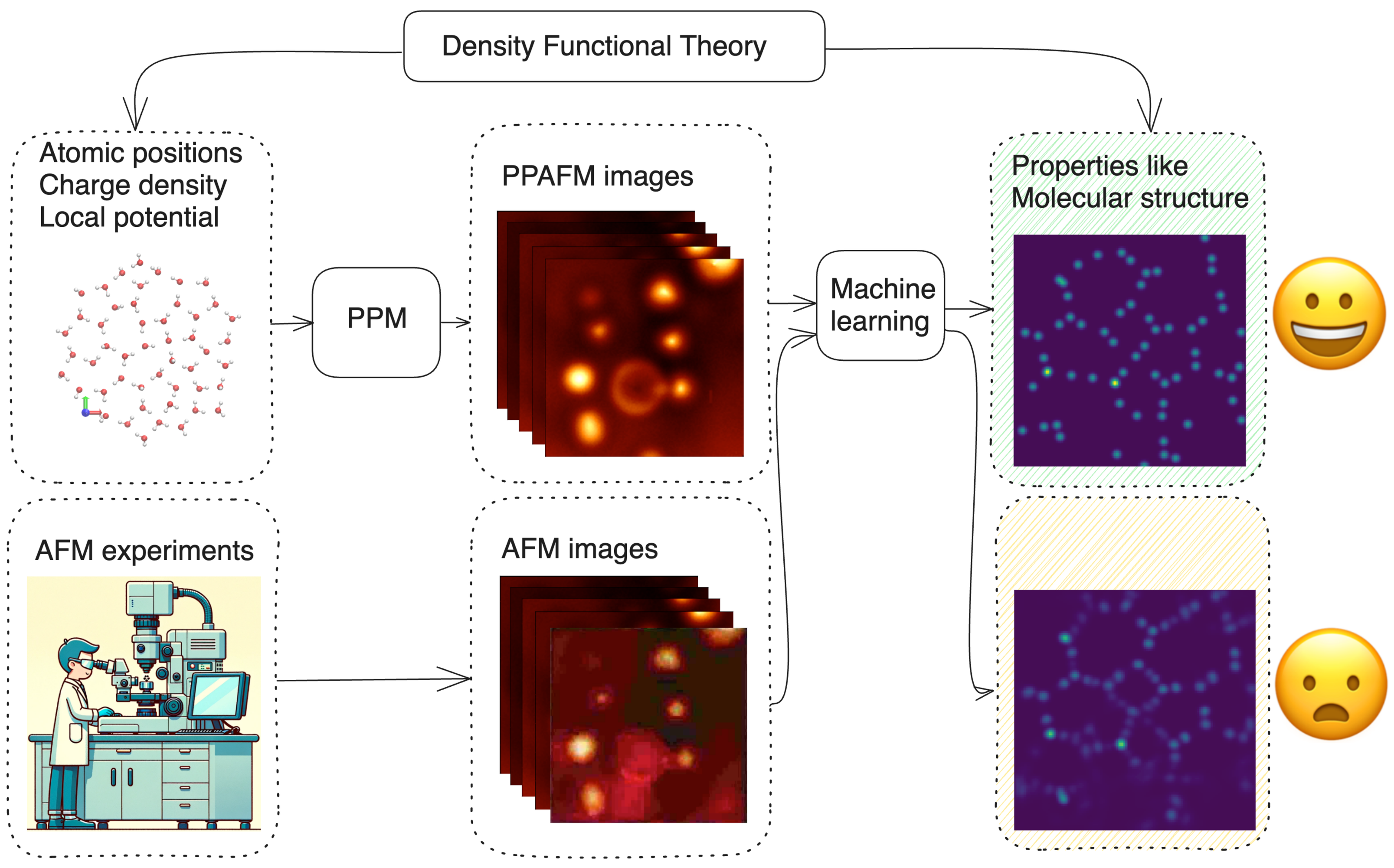


Figure 1: Integrating Machine Learning (ML) with PPAFM-generated datasets enables the prediction of properties such as molecular structures [2] and electrostatic force potential [3]. However, the performance of trained models on actual experimental AFM images often falls short compared to their performance on simulated images.

Motivation

How can we enhance the performance of ML models on experimental data?

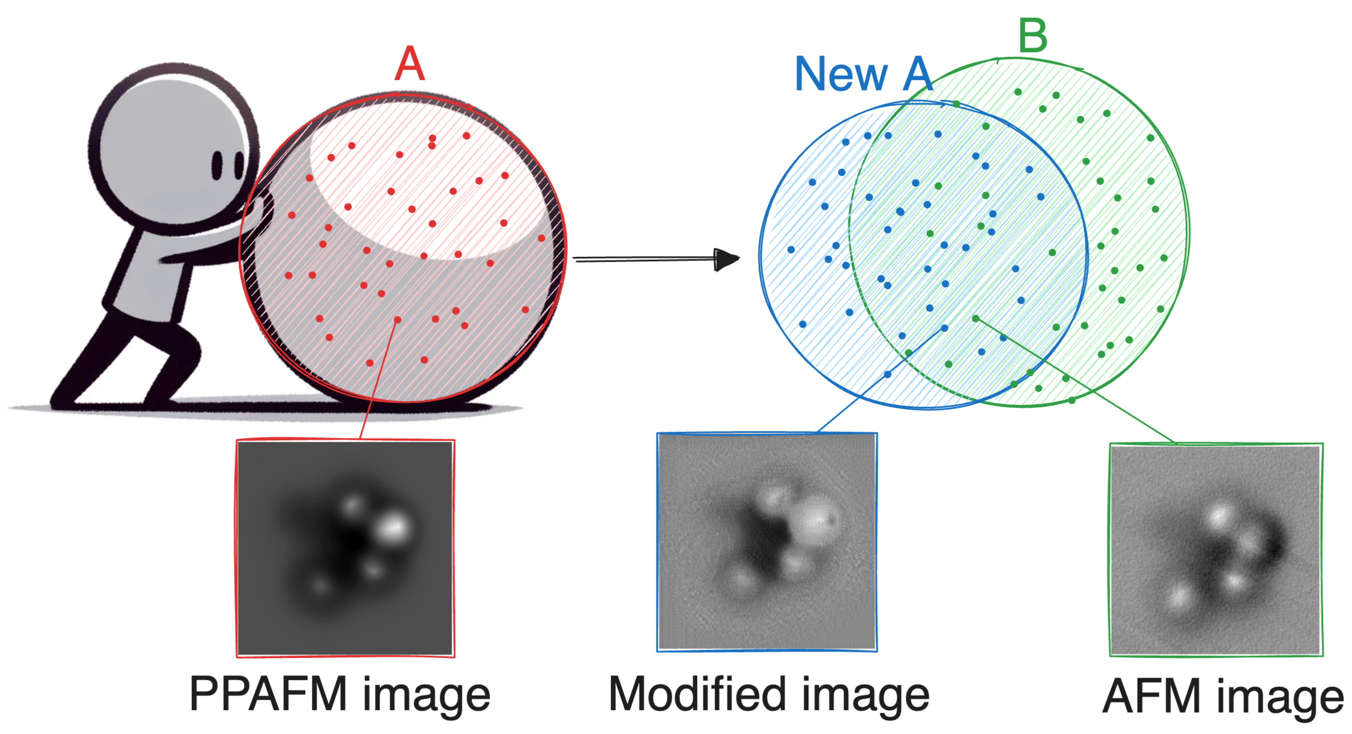


Figure 2: A potential solution is to **alter the simulated images** to more closely resemble experimental images, incorporating these modifications into the training dataset, a technique known as **Data Augmentation**. We have a large collection ofsimulated PPAFM images and a limited set of experimental AFM images.

Hypothesis

The gap between the simulation and experimental AFM images blocks the performance of machine learning models.

Question

How can we produce a large number of experimental-like AFM images from simulated ones?

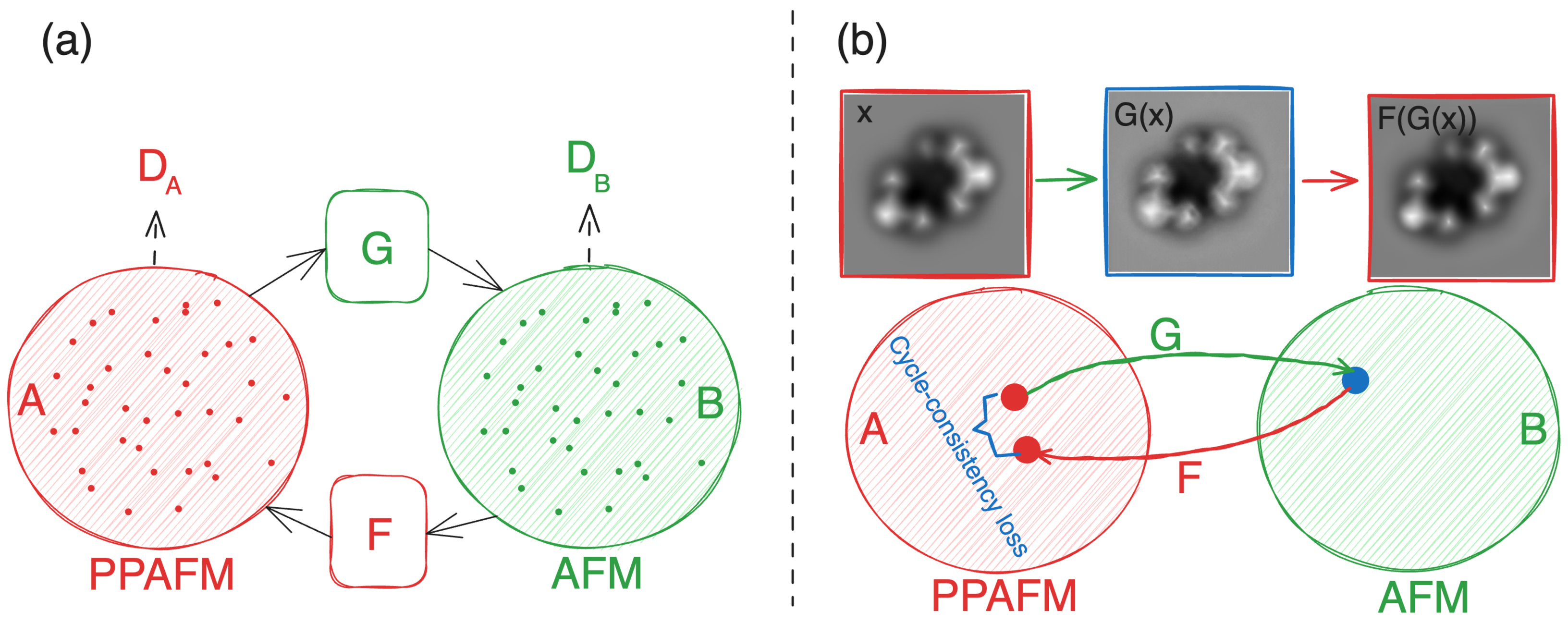


Figure 3: (a) **CycleGAN** [4] includes two mapping functions G: $A \rightarrow B$ and F: $B \rightarrow A$, and associated adversarial discriminators D_A and D_B encourages G to translate A into outputs indistinguishable from domain B, and vice versa for D_A and F. (b) **Cycle Consistency** ensures that converting from one domain to another and back again returns to the original starting point., i.e., $F(G(x)) \approx x$.

Results

The trained generator successfully introduces experimental features, such as background shadows and noise, to the PP-AFM simulation images.

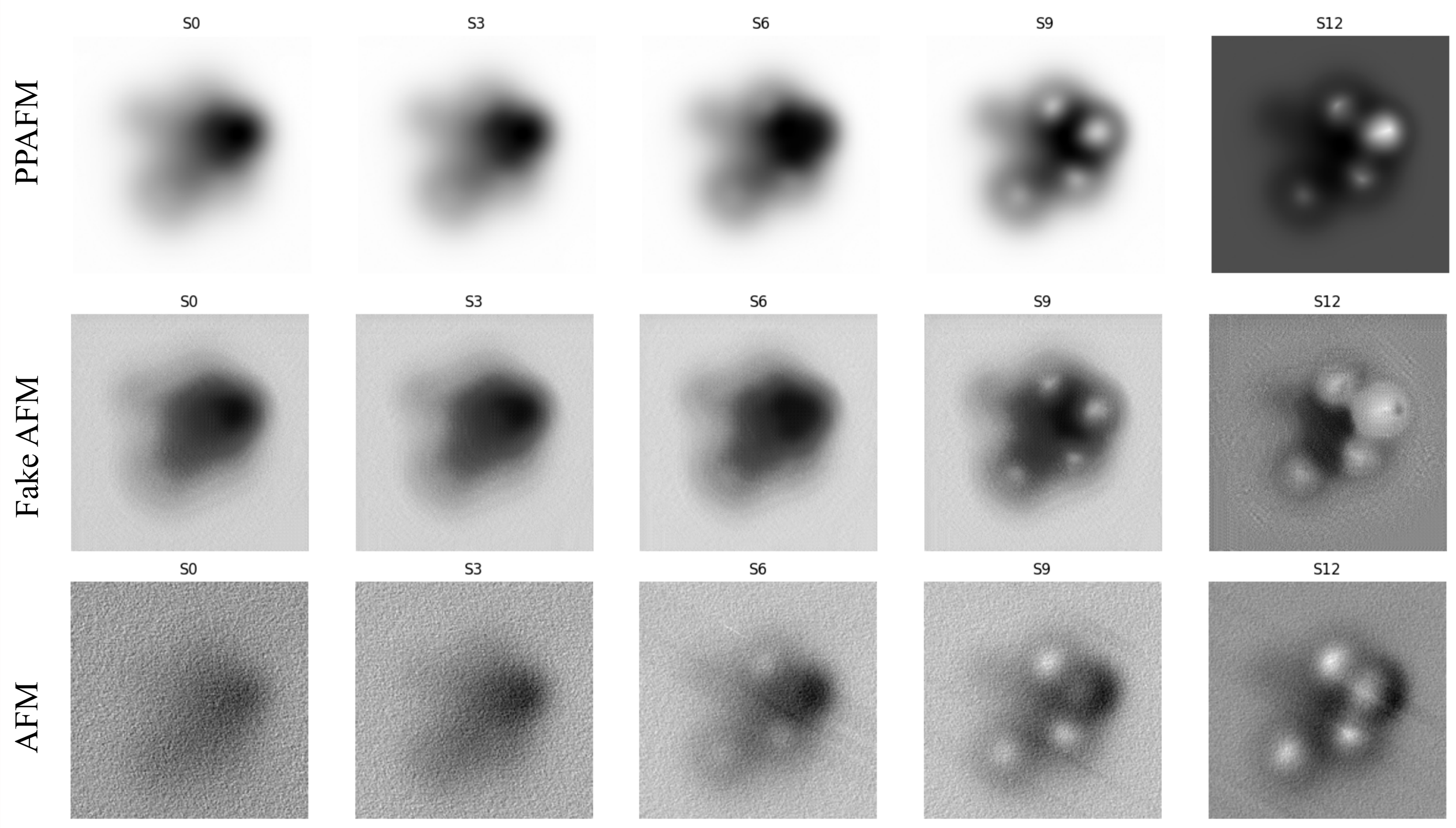


Figure 4: Depiction of water molecules on an Au surface at different heights for PPAFM simulations, generated experimental-like AFM images, and AFM images.

Future Directions

1. Design an automated method for assessing the quality of generated images.
2. Incorporate these generated images into the training datasets for machine learning models.

Acknoledgements

This work was supported by World Premier International Research Center Initiative (WPI), MEXT, Japan, and by the Academy of Finland (Projects 347319 and 346824). The authors acknowledge the computational resources provided by the Aalto Science-IT Project and CSC, Helsinki.

References

[1] P. Hapala, R. Temirov, F. S. Tautz, P. Jelínek. Origin of high-resolution iets-stm images of organic molecules with functionalized tips. *Phys. Rev. Lett.*, 113:226101, 2014.

[2] B. Alldritt, P. Hapala, N. Oinonen, F. Urtev, O. Krejčí, F. F. Canova, J. Kannala, F. Schulz, P. Liljeroth, A. S. Foster. Automated structure discovery in atomic force microscopy. *Sci. Adv.*, 6(9), 2020.

[3] N. Oinonen, C. Xu, B. Alldritt, F. F. Canova, F. Urtev, S. Cai, O. Krejčí, J. Kannala, P. Liljeroth, A. S. Foster. Electrostatic discovery atomic force microscopy. *ACS Nano*, 16(1):89–97, 2022.

[4] J.-Y. Zhu, T. Park, P. Isola, A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, *arXiv:1703.10593*, 2017.