Maskterial:
A Foundation Model for 2D Material Flake Detection from Optical Images

Jan-Lucas Uslu[1,2,4], Alexey Nekrasov[2],
Alexander Hermans[2], Bernd Beschoten[1],
Bastian Leibe[2], Lutz Waldecker[1],
Christoph Stampfer[1,3]

[1] 2nd Institute of Physics and JARA-FIT
RWTH Aachen University, Aachen, Germany

[2] Visual Computing Institute
RWTH Aachen University, Aachen, Germany

[3] Peter Grünberg Institute (PGI-9)
Forschungszentrum Jülich, Jülich, Germany

[4] Department of Physics
Stanford University, Stanford, California United States

uslu@stanford.edu

Van der Waals (VDW) heterostructures hold great promise for novel electronic and optoelectronic applications [1,2].
However, the assembly of these structures requires pristine two-dimensional (2D) material flakes.
Finding these flakes requires scanning up to 20,000 images per exfoliation to find approximately 10 suitable flakes, making manual searching very time consuming [3].

This lack of annotated data, high variability of images and extreme class imbalance within images makes training deep learning models for automated flake instance segmentation challenging.
To address this, we propose a novel model that uses physical theory to generalize from a minimal number of examples.
Our approach uses feature engineering, uncertainty estimation methods [4], and extensive synthetic pre-training [5] to inject physical inductive biases into the model.
This results in a foundation model [6] that can be fine-tuned for further downstream tasks with as few as three to five labeled examples of real data.
Extensive ablations were performed to confirm the behavior of the model and the effectiveness of the training procedures.

In the presentation, we will go over the methodology behind this approach, including techniques for incorporating domain knowledge into deep learning models and how to build and train robust models under the constraint of noisy and minimal data in physics context.

References

[1] A. K. Geim and K. S. Novoselov, Nat. Mater. 6, 183 (2007)

[2] K. S. Novoselov et al. Science 353, (2016)

[3] J.-L. Uslu et al. Machine Learning: Science and Technology 5, 015027 (2024)

[4] J. Mukhoti et al. Conference on Computer Vision and Pattern Recognition (CVPR) (2023) pp. 24384-24394

[5] G. Ros et al. Conference on Computer Vision and Pattern Recognition (CVPR) (2016) pp. 3234–3243

[6] Bommasani et al. ArXiv (2021)

Figures

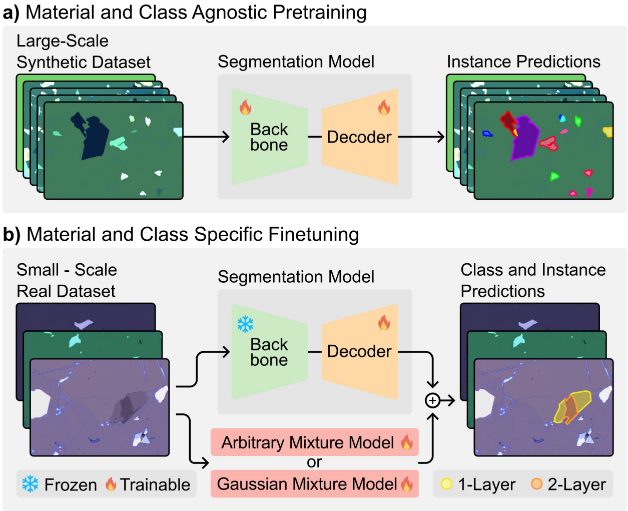

Figure 1. MaskTerial consists of two models, a segmentation model (SM) and a classification head (CH). The SM is pretrained using large-scale synthetic data from simulations tuned to by material agnostic ensuring the model only learns selected features. During finetuning only the decoder of the SM and the CH is trained on material specific data.

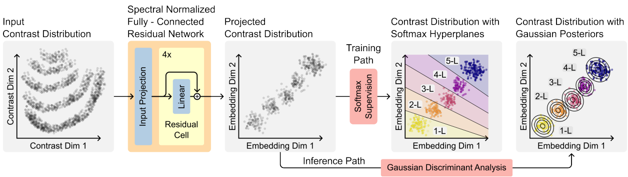
Input Contrast Distribution
Contrast Dim 2
Contrast Dim 1
Spectral Normalized Fully - Connected Residual Network
Input Projection
4x
Linear
Residual Cell
Projected Contrast Distribution
Embedding Dim 2
Embedding Dim 1
Training Path
Softmax Supervision
Contrast Distribution with Softmax Hyperplanes
5-L
4-L
3-L
2-L
1-L
Embedding Dim 2
Embedding Dim 1
Contrast Distribution with Gaussian Posteriors
5-L
4-L
3-L
2-L
1-L
Embedding Dim 2
Embedding Dim 1
Inference Path
Gaussian Discriminant Analysis

Figure 2. The Arbitrary Mixture Model (AMM) is used to quantify the uncertainty of the input data given the training data. This is done by first projecting the input contrast distribution to gaussian distributions and finally using Gaussian Discriminant Analysis (GDA) to extract the estimated uncertainties from the embedding space of the network. Actually getting a well regularized embedding space which is locally consistent for uncertainty estimation is non-trivial and requires methods such as spectral normalization and special training routines. But the final result is a typically well-conditioned embedding space from where one is able to extract meaningful distances between datapoints. This is then used to predict the classes in a physically meaningful way.