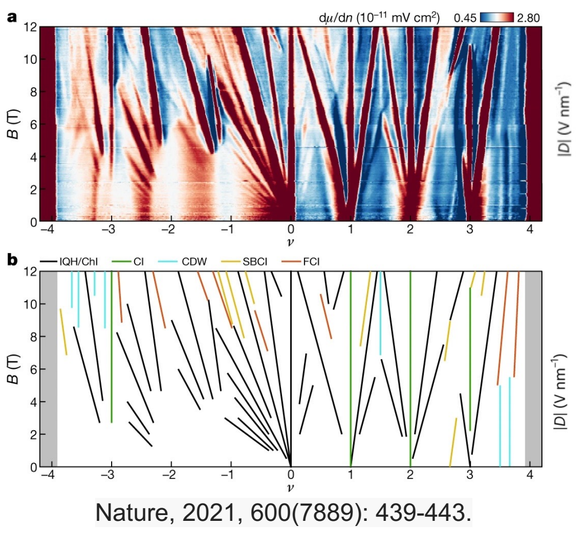
Machine Learning in Condensed Matter Data Processing

Zhaoyu Bai, Yaozhang Zhou, and Yiyang Jia

Weizmann Institute of Science, Department of Physics and Astrophysics

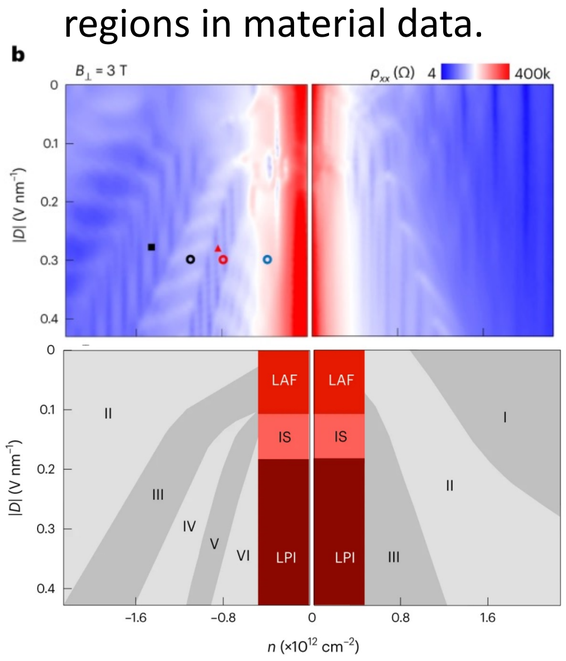
Introduction

- We try to apply Machine Learning(ML) to identify:
  - Streda Lines, representing Landau Levels in materials.
  - Phases shown as different regions in material data.



Nature, 2021, 600(7889): 439-443.

B: magnetic field. v: carrier density.



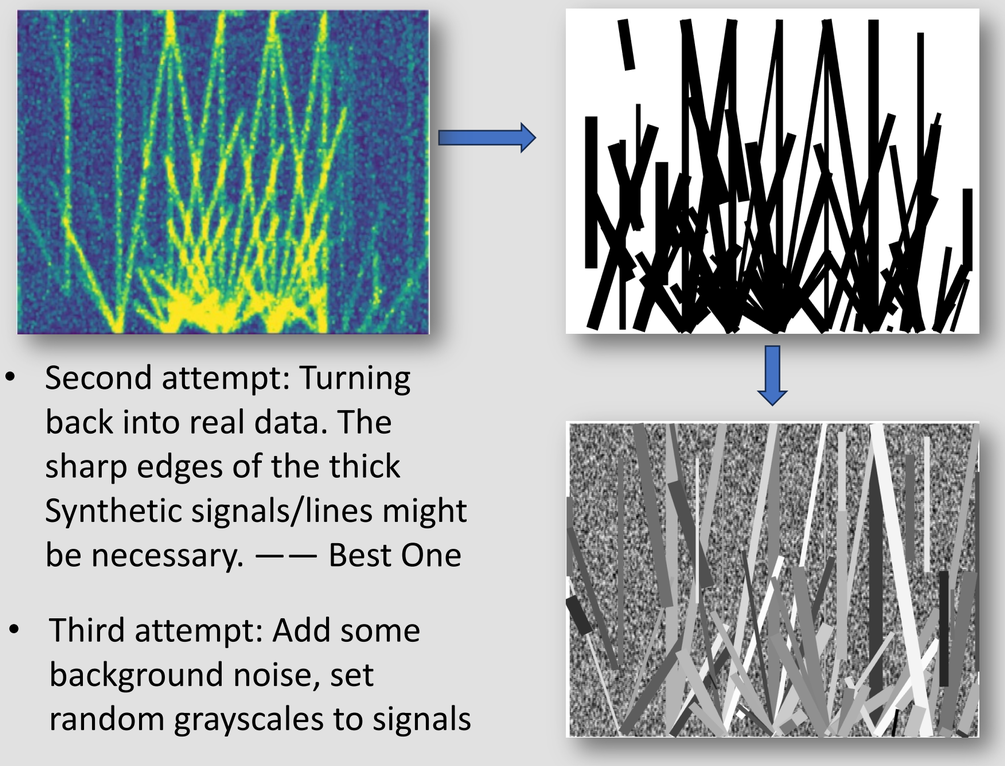
Nature Nano 19: 188–195 (2024)

D: displacement field. n: carrier density.

- ML not only reduces manual workload but also enhances the speed and accuracy of analysis, thus significantly advancing research in the field.

Finding suitable Synthetic Streda Lines

- First attempt: Add Gaussian and stripe noise into the synthetic thick lines from our immature understanding of the conditions in experiments.



- Second attempt: Turning back into real data. The sharp edges of the thick Synthetic signals/lines might be necessary. —— Best One
- Third attempt: Add some background noise, set random grayscales to signals

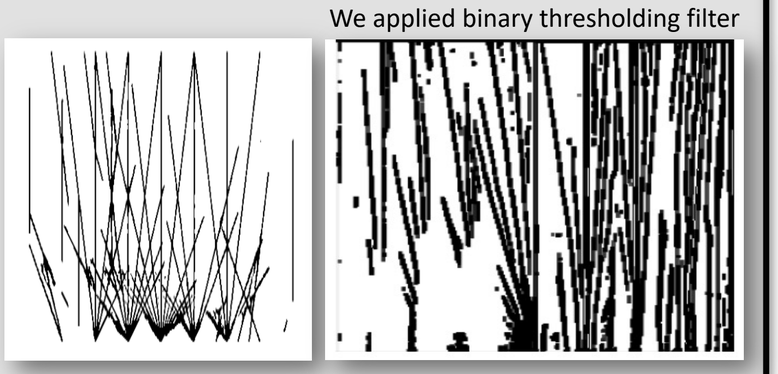
Getting Suitable synthetic data is the main difficulty.

- We learned that non-signals are not random due to various effects.
- To proceed, we need a better understanding of the non-signal background

Model Performance

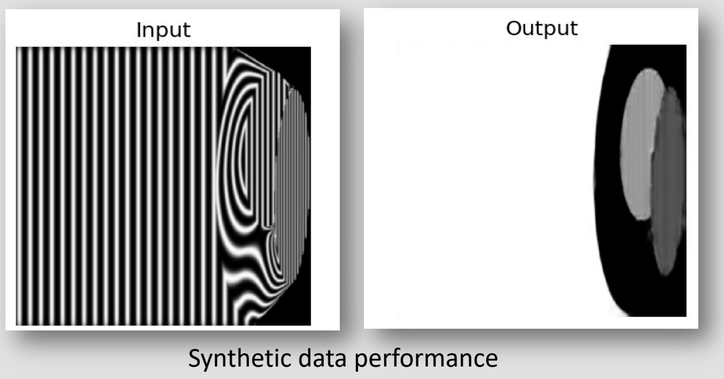
- Streda Lines:
- 1800 synthetic pictures
- Train validation ratio: 4:1
- Batch size:6
- Optimizer: Adam
- Learning rate 1e-3 3 epochs
- Learning rate 1e-4 3 epochs
- 6 epochs in total
- Hough Line detection

We applied binary thresholding filter



Synthetic data performance Data from the original paper

- Phase Segmentation:
- 500 synthetic pictures
- Train validation ratio: 4:1
- Batch size:6
- Optimizer: Adam
- Modified Loss: 3 epochs Learning rate 1e-3
- Learning rate 1e-3 5 epochs
- Learning rate 1e-3 20 epochs



Synthetic data performance

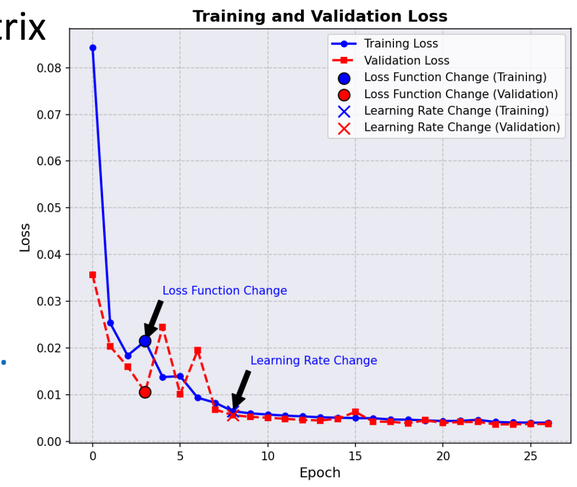
Loss Function for Phase Recognition

- Due to the bright lines in the synthetic data, ordinary MSE does not work

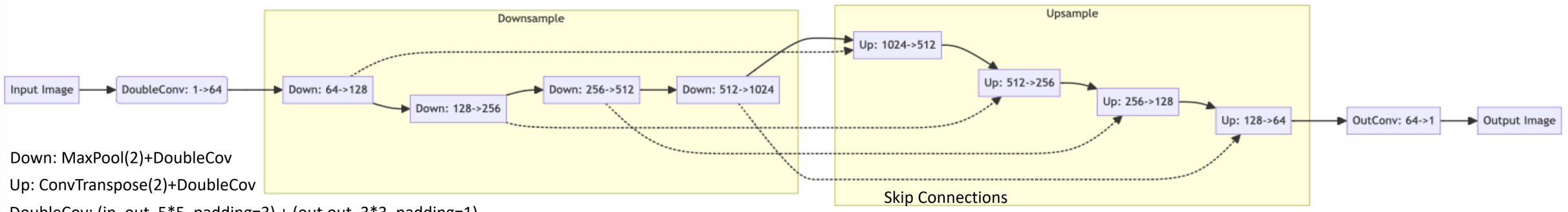
$$\mathcal{L} = \mathcal{L}_{\mathrm{MSE}} + \alpha \sum_{x,y} \mathrm{Abs}\left[\partial_x I(x,y)\right]$$

- $I(x,y)$ is the gray scale matrix
- First set α as 1. Once the white line background is removed, turn it off.

Turn-off around the epoch 3.



U-Net Model



Down: MaxPool(2)+DoubleCov

Up: ConvTranspose(2)+DoubleCov

DoubleCov: (in, out, 5*5, padding=3) + (out,out, 3*3, padding=1)

OutCov: (in, out, 5*5, padding=3) + (out,out, 3*3, padding=1)