Epitaxial Thin Film Interface Imaging with Deep Learning

Pranav Kakhandiki
School of Applied and Engineering Physics
Cornell University
Ithaca, NY 14853
pak225@cornell.edu

Yimeng Min
Department of Computer Science
Cornell University
Ithaca, NY 14853
ym499@cornell.edu

Ankit S. Disa
School of Applied and Engineering Physics
Cornell University
Ithaca, NY 14853
asd47@cornell.edu

Abstract

Complex oxide thin films exhibit unique and useful properties for electronics, energy, communications, and more. Imaging the atomic-scale structure of these films is crucial for deducing and ultimately engineering their functional behavior, but standard x-ray diffraction techniques suffer from the phase retrieval problem, which is exacerbated for nanometer sized films. Current approaches analyze crystal truncation rod (CTR) diffraction using constrained iterative algorithms to output a 3D electron density to obtain the structure. Unfortunately, state-of-the-art methodologies are typically heavily dependent on initial guesses, require high data density, and fail for thick films. Here, we propose and demonstrate a new machine learning-based phase retrieval technique for thin films – Machine Learning for Material Bragg-rod Analysis (MAMBA). MAMBA is based on a U-Net architecture that takes in the measured CTR intensity as input, and outputs the complex scattered electric field, from which the electron density $\rho(\vec{r})$ can be obtained by Fourier inversion. We summarize the promising results from MAMBA using simulated data, showing its potential for providing high-precision atomic structures of thin films beyond limitations of standard phase-retrieval techniques.

1 Introduction

When miniaturized to the atomic scale, the properties of materials can look vastly different than they do in their bulk form. This is especially true in complex transition metal oxides, which are known to have a strong interplay between their electronic and structural degrees of freedom in the system [1]. a notable example is the $LaAlO_3$/$SrTiO_3$(LAO/STO) interface, which has been found to be metallic and even superconducting, despite both materials being wide-gap insulators on their own [2]. These emergent properties are tied to distortions in the structure and charge makeup at the interface due to the strain and interfacial coupling at the interface [3]. Engineering the novel properties of complex oxide thin films could enable a new wave of technologies for energy storage, information processing and more. Key to this task is the ability to visualize the atomic-scale structure at the interface.

X-ray diffraction provides highly sensitive non-destructive, atomic-scale structural determination in crystals. While the Bragg peaks of the x-ray diffraction data are bulk sensitive, information about the interface is encoded in the scattering the peaks along the surface normal, known as crystal truncation

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

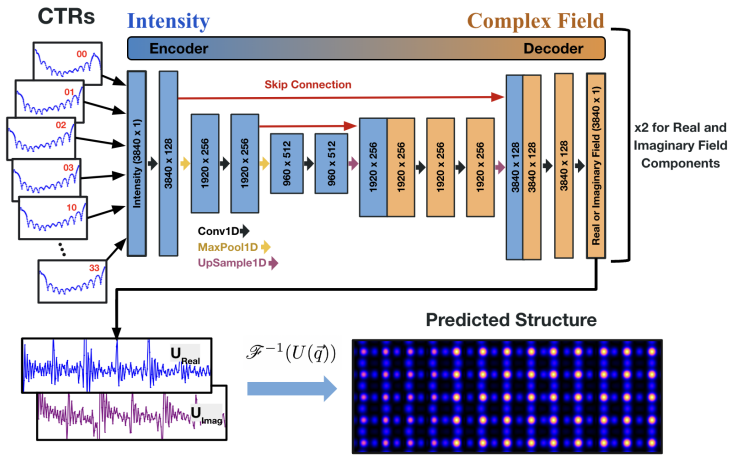

Figure 1: U-Net pipeline. The rods are flattened into one 1-dimensional array before being passed in. The U-Net uses Conv1D layers to learn features, the MaxPool1D and UpSample1D to downsample and upsample respectively. Skip connections are also employed to retain information from earlier in the network. Between layers the ReLU activation function is used. Two U-Nets are trained to predict the real and imaginary parts of the field, then an inverse Fourier transform is applied to retrieve the structure.

rods (CTRs) [4]. CTR scattering is typically orders of magnitude weaker than the Bragg diffraction, making measurement and analysis challenging. The development of high-brightness sychrotron and x-ray free electron laser sources has enabled the measurement of such subtle features. However, analysis of CTR diffraction still suffers from the well known phase retrieval problem in x-ray imaging, since experimental systems measure only the intensity and not the full scattered electric field.

One of the the state of the art phase retrieval algorithms for CTR analysis is called Coherent Rod Bragg Analysis (COBRA)[5]. However, COBRA relies on making an accurate starting guess to converge, and requires densely spaced data to converge. Thus as the film grows in size this becomes an increasingly nontrivial optimization, and it is observed that COBRA greatly struggles for films much larger than 10-12 layers.

With the recent development of robust machine learning algorithms and scalability, there have been many efforts to employ machine learning for various phase retrieval and x-ray related tasks[6–15]. However these efforts have been towards larger scale imaging of bulk crystals or x-rays from wavefront sensors, not individual atomic sensitivity of films. Furthermore, many of these models rely on physical constraints, whereas there is little prior information to deduce reasonable film perturbations.

In this paper, we use deep learning techniques to perform the phase retrieval process for an arbitrary epitaxial thin film. If the film is grown in the z-direction, this entails ultimately predicting the z-axis positions of all atoms (here we assume x-y symmetry). Occasionally, atoms from the interface and the bulk can swap, a phenomenon called charge mixing – thus the entire problem is to accurately predict atomic z-positions and charge mixing in a thin film, for a given set of CTR data. Here we demonstrate that a U-Net is able to accurately predict the missing field for a simulated 12 layer $\mathrm{LaAlO_3/SrTiO_3}$ interface, where then an inverse Fourier transform retrieves the film structure.

2 Results and Discussion

LAO/STO Simulation The idea is to train a machine learning model on many slightly-varied films that produce different CTRs, thus teaching it to predict the structure $\rho(r)$ given the corresponding intensity measurement $I(\vec{q})$. Simulations are able to match experimentally measured CTRs with great accuracy, which gives the motivation for training on simulated data. We choose the $\mathrm{LaAlO_3/SrTiO_3}$ interface, primarily because of the existing literature[2]. Here we have a bulk of $\mathrm{SrTiO_3}$ with 4 layers of $\mathrm{LaAlO_3}$ resting atop. We assume that the atomic z-positions of the top 8 layers of the $\mathrm{SrTiO_3}$ can

vary, and that the entire 4 layers of the $LaAlO_3$ film can vary. Since the film is epitaxial we assume an $x-y$ symmetry, so only the z-positions are varied. Here the positions of the A-site atoms are defined by the z-direction lattice parameter c. We vary the c_n spacing ± 0.1 Å, and the $\vec{r}$ positions of the smaller atoms $\pm 5\%$.

A phenomenon known as charge mixing can occur, where atoms residing in the same site near the interface can mix, creating mixed layers of Sr and La. Here we implement it for the 4 layers closest to the interface. This can be represented by mixing the form factors of the A-site atoms:

$$\begin{aligned} f_{mix,n}(\vec{q}) = p_n f_{Sr}(\vec{q}) + (1-p_n) f_{La}(\vec{q}), \\ p_1 + p_2 + p_3 + p_4 = 2, \text{with } p_n < 1 \end{aligned} \tag{1}$$

where layer n ranges from 1 to 4, starting with the first mixed layer, and p_n represents the percent of Sr in the n^{th} layer. $f_{mix}(\vec{q})$ then represents a pseudo-atom that is part Sr and part La. p_n is randomly varied according to the above constraint, which creates materials that have arbitrary amounts of charge mixing. Taking into account the variable atomic positions and charge mixing, the following equation accurately simulates the CTR scattering intensity of our material:

$$\delta(q_x a - 2\pi h)\delta(q_y b - 2\pi k)|F_B(\vec{q}) \sum_{-Nc}^{-1} e^{in_3 q_c c} + \sum_{0}^{N_B} F_{Bn}(\vec{q}) e^{in_3 q_c c_n} + \sum_{N_B}^{N_B+N_S} F_{Sn}(\vec{q}) e^{in_3 q_c c_n}|^2 \tag{2}$$

Where $F_{B,n}(\vec{q})$ and $F_{S,n}(\vec{q})$ are the n-dependant structure factors for each layer, calculated by replacing $f_{mix}(\vec{q})$ for the A-site atom the charge-mixed layers.

U-Net Architecture and Training U-Nets with encoder-decoder architectures have been shown to perform well on image-to-image mapping tasks, especially when the images have the same spatial coordinates [16]. Here we deploy a U-Net that predicts the E-field component that corresponds to the film scattering, $U(\vec{q})$. Thus if we predict $U(\vec{q})$ with the U-Net, an inverse Fourier transform returns the structure of the film: $\mathscr{F}^{-1}(U(\vec{q})) = p_{film}(\vec{r})$.

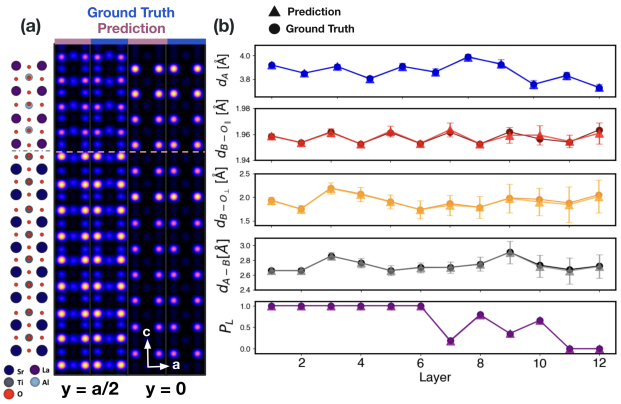

Figure 2: a) displays the predicted versus Ground Truth for $\rho(\vec{r})$. We display the slices at $q_y = 0$ and $q_y = \frac{a}{2}$ relative to the unit cell, which captures all atoms in the perovskite structure. Note the similarity in the ground truth and prediction. b) displays the extracted atomic positions for the predicted structure, compared to the ground truth structure. The error bars are the predicted error with respect to the degree of variation in the training dataset, and represent the average error in the prediction of that respective metric.

Figure 1 displays a diagram of the pipeline and the exact architecture. The CTRs are passed in as input, and for the same grid in reciprocal space, the corresponding field $U(\vec{q})$ is predicted. We flatten the input data to 1D, since the interesting information in CTRs is predominantly in the q_z direction. The rods are passed into a 1D U-Net, which has MaxPool1D layers in the encoder and UpSample1D layers in the decoder to reach back to the original space. Skip connections, which essentially transport data from earlier parts of the network untouched, as employed, as is standard practice with the U-Net. There are a total of five convolutional layers. Rather than using complex valued weights to predict $U(\vec{q})$ directly, we simply use the same architecture to train two U-Nets which predict the pair $(U_{real}(\vec{q}), U_{imag}(\vec{q}))$. Finally an inverse Fourier transform recovers the structure.

We choose the Mean Absolute Error (MAE) Loss ($\frac{1}{n}\sum_{i=1}^{n}|U_i(\vec{q}) - \hat{U}_i(\vec{q})|$) to train the model. In this case, this corresponds to the average error in the predicted versus ground truth field.

Electron Density Predictions We analyze the results of taking the inverse Fourier transform of the predicted field, returning the electron density of the film. Zero padding and Gaussian windowing

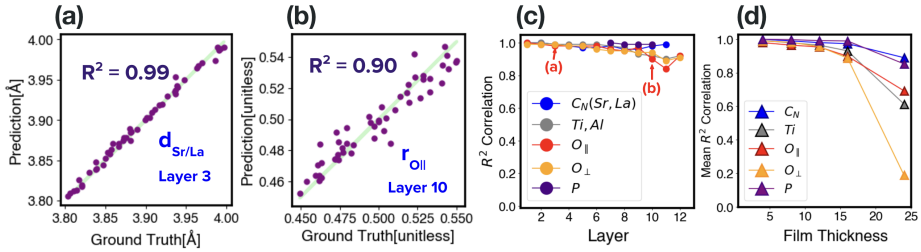

Figure 3: Quantitative analysis of MAMBA performance. a) and b) display the Predicted versus ground truth plots for two different atoms, with a sample size of N = 50 test structures. c) displays the R2 correlation coefficient for all atoms in the structure. a) and b) are pointed out here. d) displays the average R^2 coefficient for each type of atom in the ABO_3 structure as the overall film thickness is varied

were used in order to produce an image of $p(\vec{r})$ with the desired resolution. Figure 2a displays the ground truth and prediction for a randomly selected CTR. Note that the two look quite similar, demonstrating that MAMBA passes the visual test of being able to output a prediction that corresponds to a reasonable structure. Figure 2b displays the extracted atomic positions for the structures. We observe that the model is able to predict the atomic positions quite well, as well as the charge mixing parameter p.

However, since all the generated structures are a result of small atomic perturbations from their original positions, all structures are bound to look quite similar. Figure 3 shows a complete quantitative analysis. The most important metric of analysis is the R^2 value of the predicted v. ground truth plot for each atom, because this explores how sensitive the model is in detecting the small perturbations. We observe that overall all atoms are predicted quite well, with the exact R^2 values displayed in figure 3c. Note that the predictions for the c spacing are better than those for smaller atoms, likely because they have a larger impact on the CTR scattering. The predictions for the charge mixing parameter p are notably better than those of the smaller atoms as well. We also observe a slight downward trend in performance close to the top of the film.

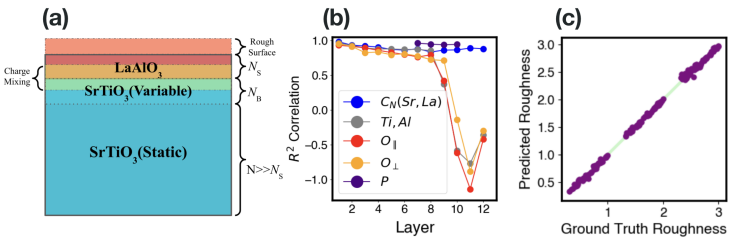

Figure 4: Performance on films with variable film thickness and rough top layer. a) shows a schematic of the simulated LAO/STO system. b) displays the overall model performance by R^2 c) shows the roughness predictions. The thickness and roughness are combined into one parameter describing the occupancy of the final layers.

We also analyze the performance if we train models for various film thicknesses. We observe near perfect performance for small 4 layer films, to acceptable performance on 16 layer films, and poorer performance on 24 layer films, especially on smaller atoms. Note that even in the 24 layer film the c spacing and p are still predicted with reasonable accuracy, and still may be of use on their own. Often in experiments the precise film thickness is unknown, and the top layer can also have a fractional occupancy, indicating a rough surface. Figure 4 displays the results of varying the overall film thickness from 11 to 13 layers, while also varying the roughness of the top layer with a random value $\in$ [0.3, 1]. The roughness is simply represented by multiplying the structure factor of the final layer by the generated fractional value. We observe that the final layer occupancy parameter is predicted quite well. However, the performance for the small atoms is significantly worse, as shown in figure 4b. The c spacing and charge mixing are still predicted with good accuracy. One important note is that the R^2 metric harshly penalizes outliers, which sometimes occurred in the 24 layer film and films with varied roughness. Thus to get a fairer metric of performance if the extracted atomic value is exceeds the region of variation by 10%, the default value is used in the calculation instead.

3 Conclusion

Phase retrieval for epitaxial thin films is necessary if one wants to unearth the exotic properties of many of the exciting quantum materials being studied. We have trained a U-Net to predict the real space electron distribution $\rho(\vec{r})$, given the CTR scattering $I(\vec{q})$. Performed on simulated data for the $LaAlO_3$/$SrTiO_3$ epitaxial interface, the model proves to perform well. We highlight that we analyze only simulated data, and while CTRs can be simulated with great accuracy, experimental data always brings more challenges. The future of this project will be in applying this to other systems and ultimately predict the structure of a real epitaxial film.

Acknowledgments and Disclosure of Funding

This project is partially supported by the Eric and Wendy Schmidt AI in Science Postdoctoral Fellowship, a Schmidt Futures program; the National Science Foundation (NSF) and the National Institute of Food and Agriculture (NIFA); the Air Force Office of Scientific Research (AFOSR); the Department of Energy; and the Toyota Research Institute (TRI)

References

[1] Elbio Dagotto. Complexity in strongly correlated electronic systems. *Science*, 309(5732): 257–262, 2005.

[2] A Ohtomo and HY Hwang. A high-mobility electron gas at the laalo3/srtio3 heterointerface. *Nature*, 427(6973):423–426, 2004.

[3] PR Willmott, SA Pauli, R Herger, CM Schlepütz, D Martoccia, BD Patterson, B Delley, R Clarke, D Kumah, C Cionca, et al. Structural basis for the conducting interface between laalo 3 and srtio 3. *Physical review letters*, 99(15):155502, 2007.

[4] Ankit S Disa, Frederick J Walker, and Charles H Ahn. High-resolution crystal truncation rod scattering: application to ultrathin layers and buried interfaces. *Advanced Materials Interfaces*, 7(6):1901772, 2020.

[5] Yizhak Yacoby, Mukhles Sowwan, Edward Stern, Julie O Cross, Dale Brewe, Ron Pindak, John Pitney, Eric M Dufresne, and Roy Clarke. Direct determination of epitaxial interface structure in gd2o3 passivation of gaas. *Nature Materials*, 1(2):99–101, 2002.

[6] Kaiqiang Wang, Li Song, Chutian Wang, Zhenbo Ren, Guangyuan Zhao, Jiazhen Dou, Jianglei Di, George Barbastathis, Renjie Zhou, Jianlin Zhao, et al. On the use of deep learning for phase recovery. *Light: Science & Applications*, 13(1):4, 2024.

[7] Logan G Wright, Tatsuhiro Onodera, Martin M Stein, Tianyu Wang, Darren T Schachter, Zoey Hu, and Peter L McMahon. Deep physical neural networks trained with backpropagation. *Nature*, 601(7894):549–555, 2022.

[8] Di Chen, Yiwei Bai, Sebastian Ament, Wenting Zhao, Dan Guevarra, Lan Zhou, Bart Selman, R Bruce van Dover, John M Gregoire, and Carla P Gomes. Automating crystal-structure phase mapping by combining deep learning with constraint reasoning. *Nature Machine Intelligence*, 3 (9):812–822, 2021.

[9] Guohao Ju, Xin Qi, Hongcai Ma, and Changxiang Yan. Feature-based phase retrieval wavefront sensing approach using machine learning. *Optics express*, 26(24):31767–31783, 2018.

[10] Yue Wu, Lin Zhang, Siqi Guo, Limin Zhang, Feng Gao, Mengyu Jia, and Zhongxing Zhou. Enhanced phase retrieval via deep concatenation networks for in-line x-ray phase contrast imaging. *Physica Medica*, 95:41–49, 2022.

[11] Ohsung Oh, Youngju Kim, Daeseung Kim, Daniel S Hussey, and Seung Wook Lee. Phase retrieval based on deep learning in grating interferometer. *Scientific Reports*, 12(1):6739, 2022.

[12] Rujia Li, Giancarlo Pedrini, Zhengzhong Huang, Stephan Reichelt, and Liangcai Cao. Physics-enhanced neural network for phase retrieval from two diffraction patterns. *Optics Express*, 30 (18):32680–32692, 2022.

[13] Yuki Nagahama. Phase retrieval using hologram transformation with u-net in digital holography. *Optics Continuum*, 1(7):1506–1515, 2022.

[14] Kenan Li, Guanqun Zhou, Yanwei Liu, Juhao Wu, Ming-fu Lin, Xinxin Cheng, Alberto A Lutman, Matthew Seaberg, Howard Smith, Pranav A Kakhandiki, et al. Prediction on x-ray output of free electron laser based on artificial neural networks. *Nature Communications*, 14(1): 7183, 2023.

[15] Yunzhe Li, Yujia Xue, and Lei Tian. Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media. *Optica*, 5(10):1181–1190, 2018.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pages 234–241. Springer, 2015.

[17] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

4 Supplementary Materials

4.1 Preprocessing

The rods passed into the network are the 16 rods corresponding the region from 00 to 33. q_z ranges from 0 to 6 with Δq_z = 0.025, so our region of reciprocal space is $q_x, q_y \in [0, 3], q_z \in [0, 6]$. This makes each rod of dimension 240, for a total of 3840 points being passed into the network as one flattened array. For the input, due to the many orders of magnitude difference between the Bragg peaks and the in-between fringes, we take the log of the data, where $X_{train} = log(I(\vec{q}))$.

The data is then normalized to mean 0 and standard deviation 1. For the output data, since U_{real} and U_{imag} can have negative values, we apply a transformation where each point is mapped to its z-score relative to all other CTRs simulated in the dataset $\mathcal{D}$. So for a certain point in reciprocal space at (q_x, q_y, q_z), for U_{real}, the transformation looks like: $y^i_{train,real} = (U^i_{real} - \mu_i)/\sigma_i$ where the superscripts and subscripts i correspond to an arbitrary (q_x, q_y, q_z) point, and μ_i and σ_i are calculated using the training dataset $\mathcal{D}$. The process is identical for the imaginary component.

4.2 Training

To train the model, we use a batch size of 32, a learning rate $\alpha = 0.001$, and momentum decay factors $\beta_1 = 0.9, \beta_2 = 0.999$. To optimize the loss function, we use the Adam optimizer[17]. 10^5 CTRs are generated for training on, for which 10% is used for validation. The model is trained for 30 epochs, which is when the validation curve begins to plateau, ensuring the model will not over fit. Each of the two models took $\sim$ 11 hours to train.

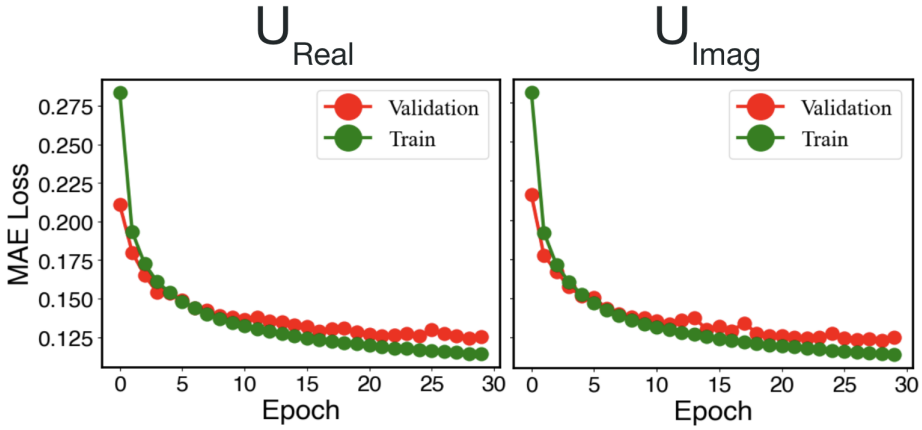

Figure 5: MAE Loss over epochs. The model is trained for 30 epochs, and the best validation error is taken as the final model.

4.3 Field Predictions

For the same electron density map in figure 2, we also plot the corresponding $U(\vec{q})$ prediction for the 00 rod. Note that the predicted field and ground truth field line up quite well, which leads to a good prediction post inverse Fourier transform. We also observe that the random field does not line up with the ground truth field, demonstrating the power of the model to accurately predict the field $U(\vec{q})$.

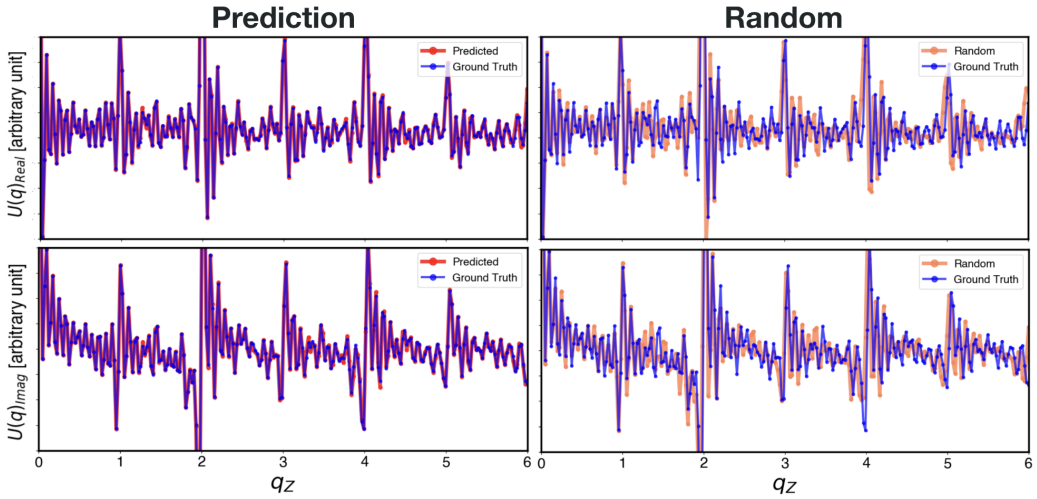

Figure 6: The plots on the left compare the prediction to the ground truth field, $U(\vec{q})$. This field corresponds to the electron density plot in figure 2. For context, on the right we also display the ground truth field against a randomly sampled different ground truth field from the dataset, to emphasize the quality of the prediction. The 00 rod prediction is displayed.