

<https://doi.org/10.1038/s42005-025-02152-8>

Physics-guided deep learning strategy for 2D structure reconstruction from diffraction patterns



Rong Fu^{1,2,3}, Tianhao Su^{1,3}, Musen Li^{1,2}✉, Yue Wu¹, Runhai Ouyang¹, Danica Solina², Michael Cortie², Tongyi Zhang¹, Shunbo Hu¹ & Zhongming Ren¹✉

Two-dimensional (2D) materials have garnered significant attention due to their tunable electronic and optical properties and exceptional mechanical performance. Reconstructing 2D structures from diffraction patterns without prior assumptions or comprehensive knowledge is challenging, especially for heterogeneous stacking and quantum 2D materials. Here, we introduce DD2D (diffraction pattern deep-reconstruction 2D structures), a physics-guided deep learning method that predicts 2D structures directly from diffraction patterns. DD2D employs a twin-tower framework, integrating a crystallographic geometric encoder and a site texture encoder, and uses a self-attention mechanism to identify intrinsic correlations in physical information and corresponding areas in the diffraction pattern. The results demonstrate high anti-interference, robust recognition capabilities, reliable interpretability, and prediction accuracy of up to 99.0%, highlighting its potential for future 2D materials discoveries.

The success of graphene in fundamental studies and applications has attracted much interest in exploring two-dimensional (2D) materials over the past decade^{1,2}. Various classes of 2D materials with a wide range of elemental compositions, have been discovered, exhibiting rich internal degrees of freedom and tunable physicochemical properties^{2,3}. The unique quantum effects and boundary conditions due to subnanoscale thickness result in altered electronic and optical properties within the 2D plane, as well as enhanced mechanical properties^{4–6}. 2D-layered materials are considered promising candidates for next-generation devices and show potential in fields such as quantum materials, sensors, catalysts, and energy storage^{6–8}.

Due to the lack of periodicity boundary condition in the out-of-plane direction, determining stacked structures of 2D crystals is more complex than that of bulk three-dimensional (3D) materials^{7–10}. Comprehensive methods, including diffraction techniques and structural models, are required to unveil the full morphology of 2D components and structures. Compared to 3D crystals, the diffraction peaks or spots captured from non-periodic direction of 2D structures are relatively weak and indistinguishable. However, the diffraction patterns provide critical out-of-plane structural alignment information necessary for reconstructing 2D structures^{11–13}, notably in heterogeneous stacking and complex quantum materials. Modeling methods based on *ab initio* Monte Carlo simulation, molecular

dynamics simulation, and refinement techniques (like the Rietveld method) have been proposed to support determining geometric structures of 2D crystals^{14–16}. However, initial structural guesses for above models are required crucially, which would be costly, time-consuming, and dependent on specialized knowledge¹⁷.

With the development of computational science, the interdisciplinary integration of AI and materials science has opened new avenues for developing functional materials^{18–23}. Developing machine learning (ML) program to assist in determining and reconstructing the unknown components will advance the exploration and industrialization of 2D materials. ML based studies have been introduced into the 2D materials community since 2018²⁴, while predicting structures from diffraction data remains unexplored. The challenge lies in developing an efficient framework with encoding algorithms that ensure the trained model deeply learns the structural features and is grounded in physicochemical knowledge. Here we propose a deep contrastive method called DD2D (diffraction pattern deep-reconstruction 2D structures), which employs a self-attention mechanism to autonomously learn the correlations between diffraction patterns and 2D material structures. Thus, DD2D can infer 2D structures from spectra without the need for initial hypotheses and relative knowledge, effectively avoiding the introduction of manual errors. The results demonstrate the potential for practical applications, as DD2D successfully recognizes

¹State Key Laboratory of Advanced Special Steel, School of Materials Science and Engineering & Materials Genome Institute, Shanghai University, Shanghai, China. ²School of Mathematical and Physical Sciences, University of Technology Sydney, New South Wales, Sydney, Australia. ³These authors contributed equally: Rong Fu, Tianhao Su. ✉e-mail: musenli@shu.edu.cn; zmren@shu.edu.cn

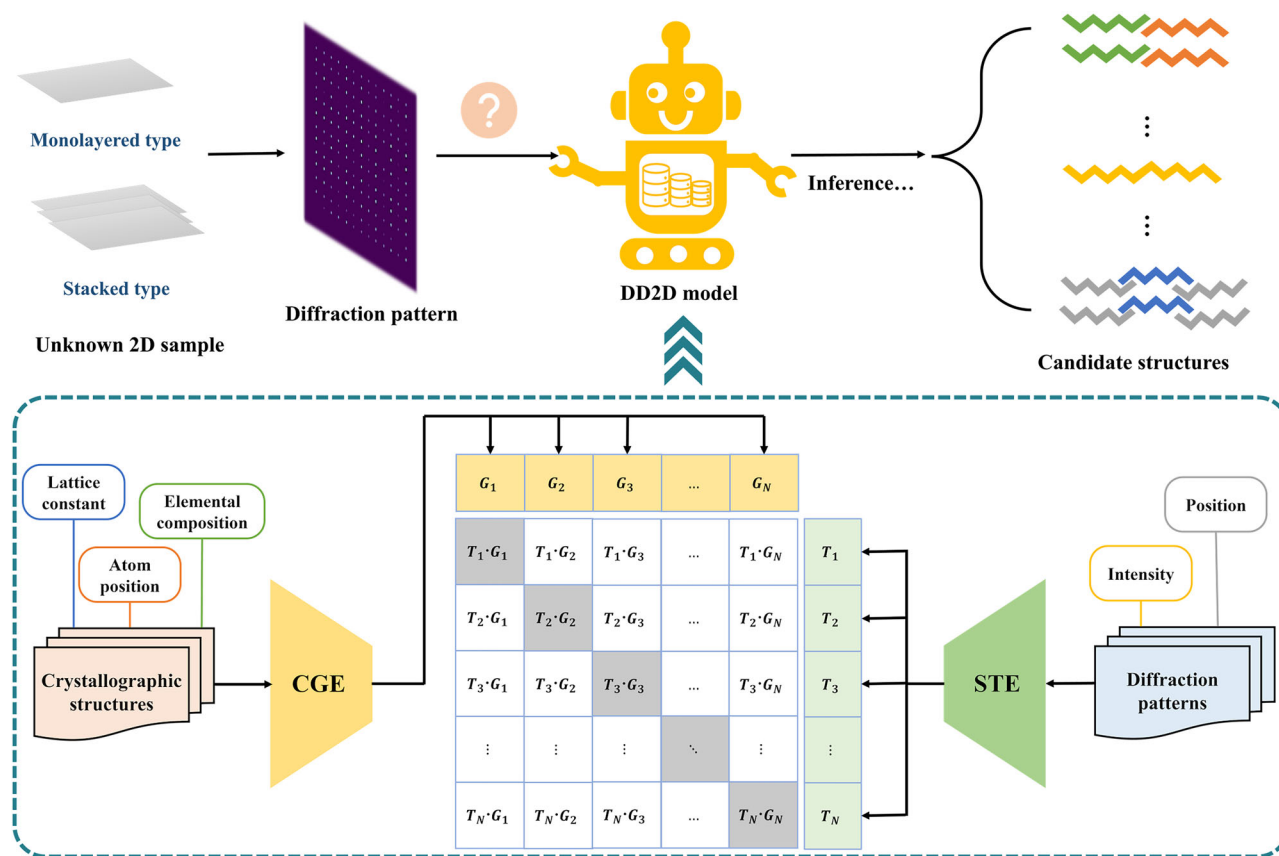


Fig. 1 | Overview of DD2D workflow. DD2D makes structure predictions based on the given diffraction patterns of unknown 2D materials. The training process of DD2D, as displayed in the dashed frame, includes the CGE and STE encoders, which process the crystal structures and diffraction image information, respectively. The

feature matrix given by the two encoders shows the similarity matrix of the comparative twin-tower model, representing the attributes and positions of different atoms. G_i is the crystallographic geometric representation (G_i) and T_i is the site texture representation.

underlying physical information and displays high predictive accuracy, robust generalizability, and reliable interpretability.

Results

The model training process, as illustrated in Fig. 1, introduces a contrastive twin-tower framework for the DD2D, which includes two transformer-based deep learning encoders. The crystallographic geometric encoder (CGE) encodes the atomic structure information, including lattice constants, atomic positions, and elemental composition, converting text sequences into a fixed length embedding that encapsulates the semantic features of 2D crystals. The input text is tokenized and encoded by the CGE, with a multi-layer self-attention mechanism employed to capture semantic dependencies among different tokens. On the other hand, the site texture encoder (STE) encodes intensity and position information extracted from diffraction patterns. Specifically, it recognizes relationships between various patches within the patterns through a multi-layer self-attention mechanism, mapping features via fully connected layers to output a fixed-length vector representing the visual characteristics of the diffraction patterns. To reduce the risk of overfitting to irrelevant regions or noise, specialized tokens target specific regions in the diffraction pattern, enabling the model to focus on key areas associated with structural information. Meanwhile, regularization and an early-stopping mechanism were applied during training to further prevent overfitting. As a result, this design bridges 2D atomic structures with their corresponding diffraction patterns, enabling direct prediction from experimental data to material structures.

The self-attention mechanism (see Eq. (3) in the Methods section) enables the model to autonomously analyze and weight substructures of the crystals. The study of attention score visualization, as shown in Fig. 2, revealed the correlation between physical information in layer 1 (the initial

transformer layer) and layer 4 (the final transformer layer) of the CGE, alongside the corresponding 2D structural details and deeply encoded features. Specifically, we sorted the lattice parameters of 2D structures, as labeled in the figure. The geometric structure of a 2D material can be considered a layered arrangement from a bulk crystal, where the c-axis contributes negligibly to the diffraction pattern due to the absence of periodic order in that direction. The length of the b-axis is generally strongly correlated with the shorter a-axis, exhibiting robust isotropic scaling constrained by the symmetry. Meanwhile, the shortest a-axis is pivotal in the diffraction pattern, providing essential information about the [100] surface of 2D crystals. As depicted in Fig. 2a–d, layer 1 visualizes the distribution of weights for crystal structure information autonomously learned across four attention heads. Since the b-axis parameter garners attention from the first to third heads, the fourth head consequently focus on a-axis parameter, which is encoded in the first sequence position of the crystal. The results clearly demonstrate that the mechanism of attention weight allocation aligns with crystallographic principles and is highly sensitive to the geometric structure of 2D materials, enabling the model to identify and learn from the most informative portions.

Interpretability is a crucial property for a neural network model, which refers to the ability to explain how a model makes decisions based on input data, ensuring that its high accuracy performance comes from extracting discriminative physical features and provides insights into the physical principles the model has learned, particularly in complex deep learning studies. As mentioned above, the four attention heads in the initial transformer layer have independently learned distinct aspects of geometric physical information, thereby collaboratively accomplishing the task of investigating and encoding all physical information. By comparing the weight distributions of layers 1 and 4, it was observed that the CGE

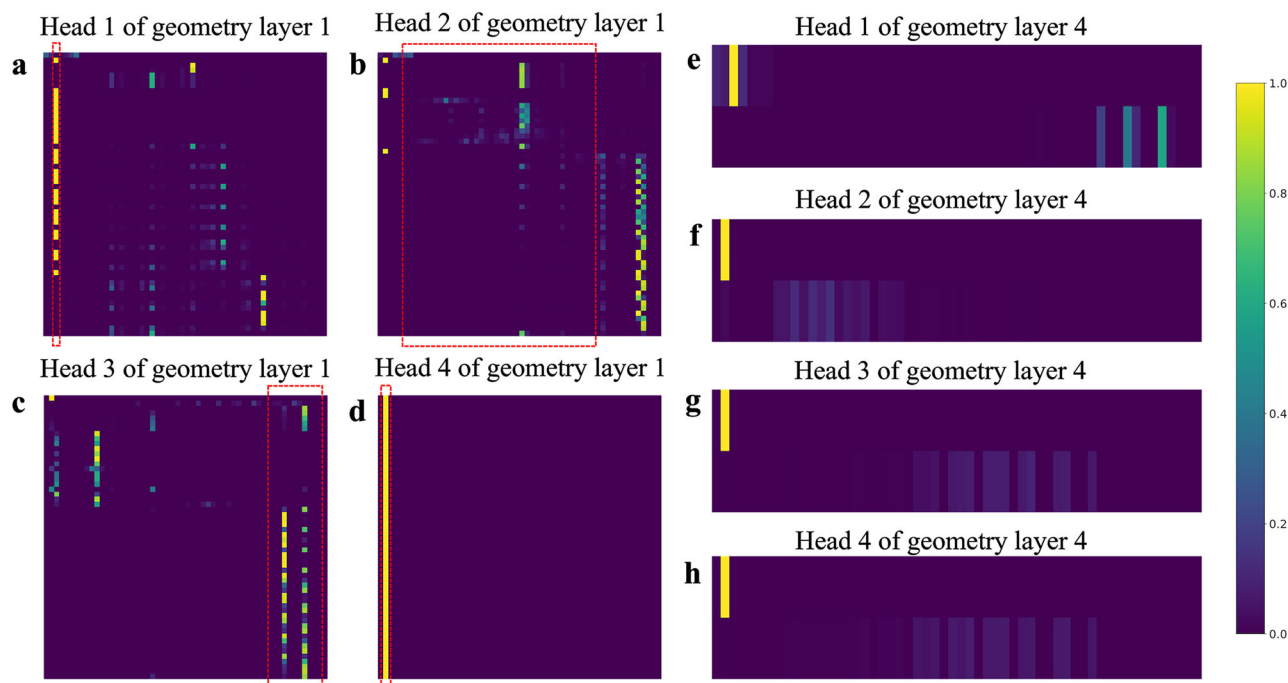


Fig. 2 | Attention visualization in the crystallographic geometric encoder block. **a–d** Heatmaps for geometry layer 1 present the inter-element attention scores within the crystal structure sequences across various attention heads in the first self-attention layer. The red dashed frames in patterns a to d represent the b-axis of lattice parameters, atomic positions, element information, and the a-axis of lattice

parameters, respectively. **e–h** Heatmaps for geometry layer 4 highlight key points in the heatmap of weights for different attention heads, associated with the two tokens [CLS_star, CLS_end] extracted from the last self-attention layer. For the [CLS_star] token at position 0 in each sequence, it is required to focus exclusively on the sequence positions corresponding to the lattice constants.

Table 1 | Accuracy of the DD2D trained on unperturbed and perturbed datasets

Attempts Dataset types	1	3	5	10
Unperturbed	62.0%	92.0%	95.0%	99.0%
Perturbed	55.0%	83.0%	94.0%	99.0%

effectively connects the information between the initial structure embedding sequence and the final deeply encoded sequence. As shown in Fig. 2e–h, while considering all relevant information, the key parameter (a-axis) simultaneously receives the maximum attention weight in the final transformer layer of the CGE, highlighting the interpretability of the model framework.

To evaluate the accuracy and anti-interference capabilities of the trained model, we developed a dataset of perturbed crystal structures using Ab Initio Molecular Dynamics (AIMD) (see Methods section), effectively simulating the noise caused by atomic thermal motion or data acquisition processes. The unperturbed test case was randomly split into training and testing datasets, ensuring the generalization ability of the trained model and preventing data leakage. Table 1 displays the accuracy performance of the model, which was independently trained on unperturbed and perturbed datasets, respectively. The compared results demonstrate that the model maintains reasonable recognition capability even when trained on datasets with reduced order, suggesting its adaptability and robustness in the face of varying degrees of data perturbation. Additionally, it is notable that the model achieved 99% accuracy when comparing the reference structure with ten predicted candidates, demonstrating a powerful recognition performance.

In addition to its high accuracy performance, we explored the unexpected predictive structures generated by the trained model to determine if the model has truly learned the underlying correlations between physical parameters (including planer symmetries and stacking orders). Briefly, the

capability of model in recognizing specific physical information can be evaluated by comparing the similarity of variables among those mismatch outputs. We analyzed the symmetry operation matrices related to the point groups within mismatched outputs for an identical diffraction pattern, utilizing hierarchical clustering²⁵ to evaluate their similarities, as depicted in Fig. 3. The mismatched structures can be fundamentally categorized into five distinct groups, each exhibiting varying degrees of successive correlation, as illustrated in Fig. 3a. Furthermore, Fig. 3b displays the distribution of point groups among predicted structures, with candidate numbers ranging from narrow to medium and wide. For example, when analyzing the prediction task targeting the *mm2* point group, the narrow range results suggest the point groups include *mm2* along with neighboring groups such as -1 , *m*, and $2/m$. If extending the predictions to the medium range, the result still includes structures that feature the target group alongside congeneric group $4/mmm$, or adjacent groups including -1 , 2 , $2/m$, and *m*. When predictions are expanded to a wide candidate number range, a subclass point group (-6) emerges, yet the composition predominantly includes target and neighboring groups. Additionally, this phenomenon is particularly pronounced in the prediction task targeting the $2/m$ point group, where results include the target along with congeneric groups, exhibiting the accuracy of the predictions. As a result, within the designed contrastive learning framework, the mismatched outputs exhibit a clear preference for certain specific symmetry types, which indicates that the trained model can recognize physical characteristics and has effectively learned the symmetrical similarities of 2D material samples in the latent space through texture analysis.

Furthermore, we tested a practical example of applying the DD2D to directly reconstruct the 2D crystal structure from a diffraction pattern [<https://github.com/SuthPhy2Ai/DD2D>], as presented in Fig. 4. The results show the ten candidates generated by the model including the target 2D crystal PrSe_3 . By comparing the atomic compositions of the mismatch candidate structures, we found that their anions are composed of elements from the VIA group (Se, O, S, Te), while their cations include rare earth metals (Sm, La, Pr, Eu) as well as transition metals (Ag, Ti). The regular

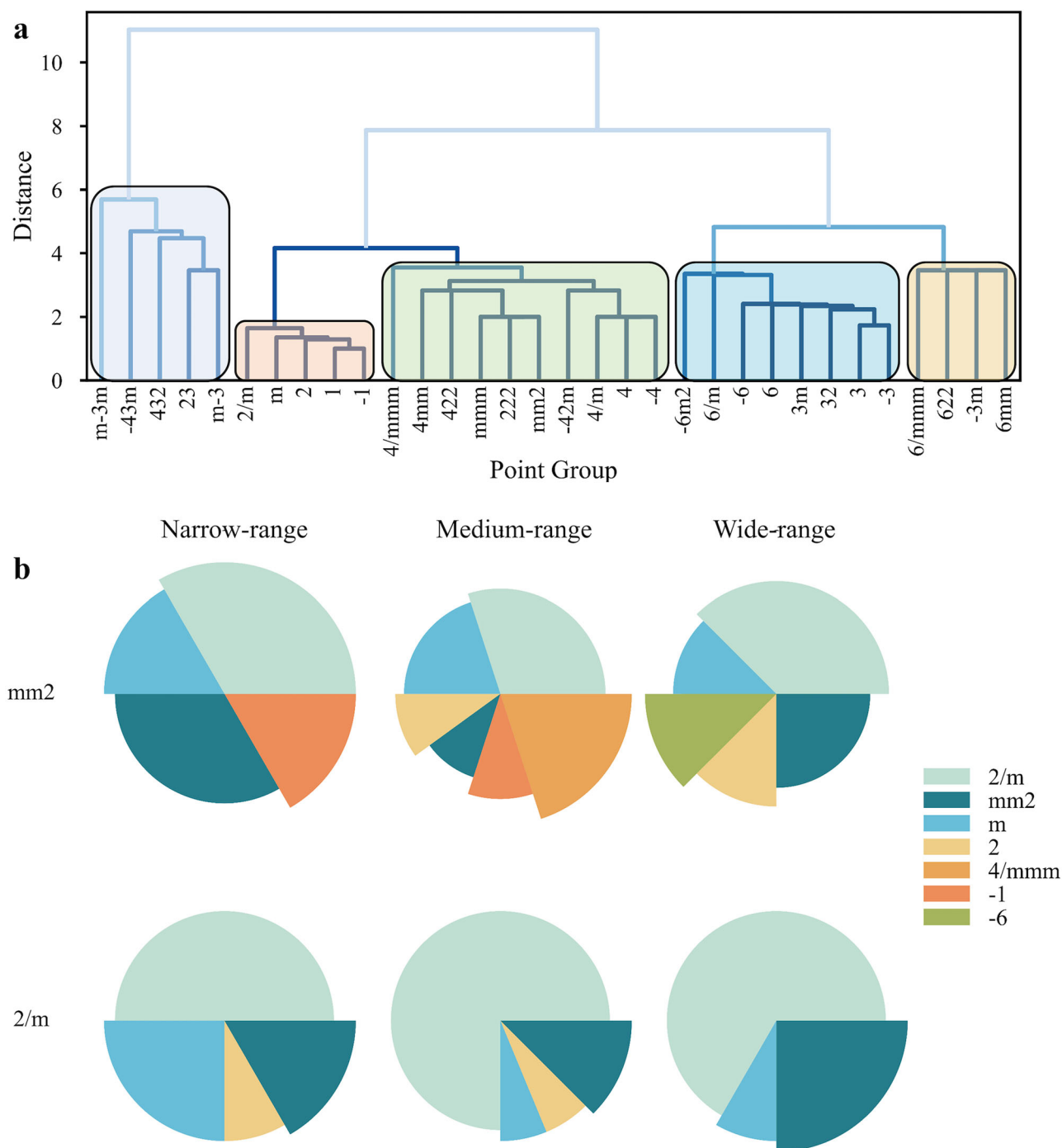


Fig. 3 | Analysis of predictions point groups. **a**, Dendrogram that represents the hierarchical clustering of point groups, showing how they merge based on their similarity. Squares of different colors identify distinct subgroups of similar point groups. **b**, Comparison of point group predictions across different ranges is shown in

the charts (left-side annotations indicate target point group symbols mm2 and 2/m). The charts use different colors to represent the point group symbols of 2D crystal structures, and the narrow-range, medium-range, and wide-range predictions correspond to attempts 1, 3, and 5, respectively.

phenomenon of element composition suggests that the self-attentional encoding mode successfully recognized intrinsic physical properties such as symmetry and elemental categories, which further demonstrates that the DD2D has learned the deep correlations between various structural information during the training process. Meanwhile, the model successfully provided the complete morphology of the generated 2D crystal materials, including the complex layered structures in an out-of-plane direction, based on a two-dimensional diffraction pattern. In other words, the DD2D can be applied not only to monolayer 2D materials but also to predicting functional heterogeneous structures, such as van der Waals 2D materials.

Discussion

We have conducted a comprehensive investigation of the DD2D, which is based on a contrastive learning dual-tower framework with CGE and STE encoders. The results demonstrate that the weight learning of various structural information by the self-attention mechanism aligns well with physical principles, and the generated candidate structures exhibit similarity in terms of symmetry and elemental composition. The DD2D has learned the underlying correlations of various physical information and demonstrated reliable interpretability, ensuring that the output reference structures are grounded in physical knowledge rather than merely algorithmic numerical

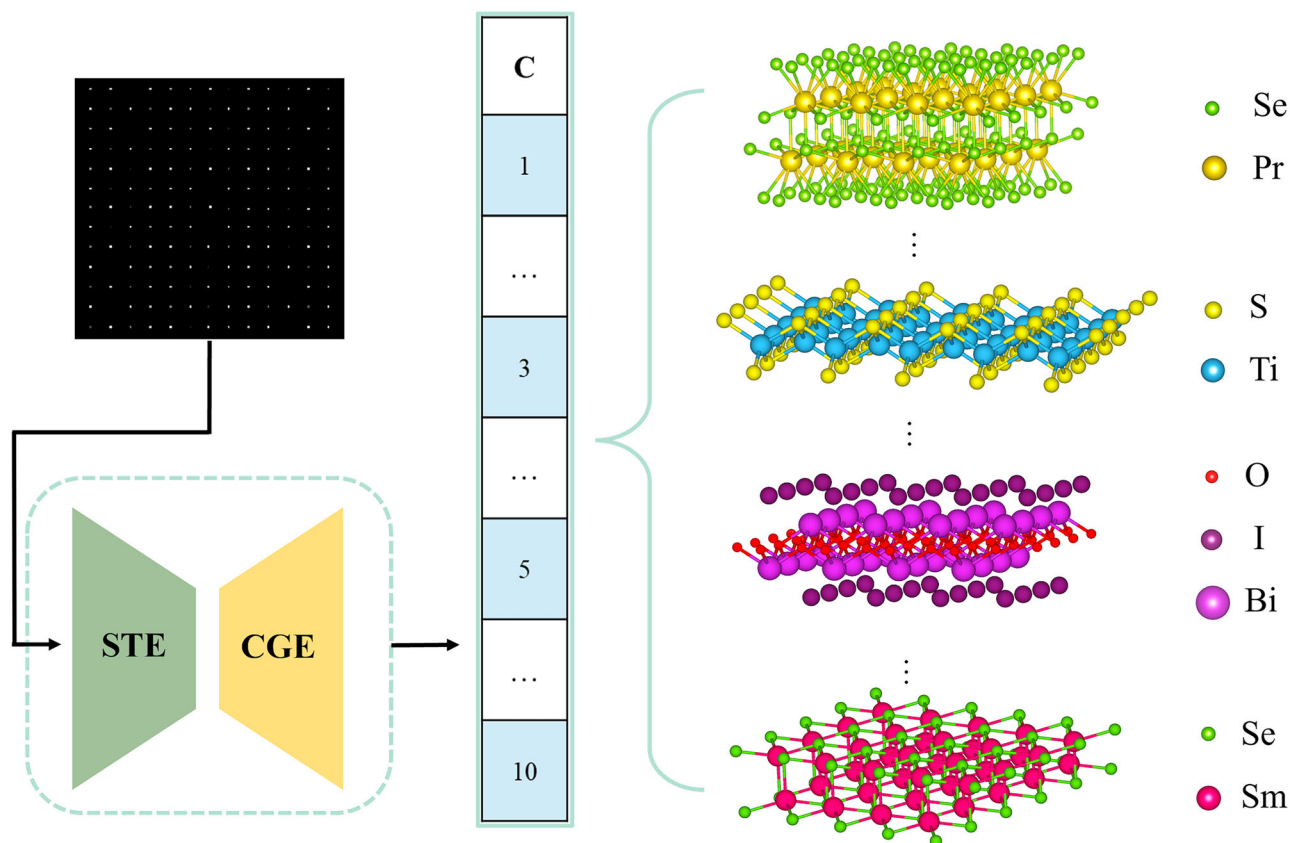


Fig. 4 | Practical test of DD2D prediction capability. A simulated diffraction pattern is recognized and learned by DD2D (left), and ten prediction attempts are provided by DD2D. For the specific case of the diffraction pattern of PrSe_3 , the

predicted structures contain the same groups of anions and are listed on the right. They include PrSe_3 , TiS , BiI , and SmSe .

fitting. Consequently, the high accuracy (up to 99.0%) and anti-interference performances presented by the model are reliable and trustworthy. The principle of most AI models is a ‘black box,’ which relies on statistical fitting and regression of large amounts of input data, often leading to uncertain predictions outside the training set. Thus, interpretable models rooted in physical information instill reliability in predictions and reveal unexpected correlations that may provide scientific insights into materials science.

Furthermore, it is important to highlight the long-standing challenge in the interdisciplinary integration of machine learning and materials science, specifically the diversity of large datasets and its impact on the generalizability of models, particularly when predicting unknown 2D material structures. In the DD2D, the largest publicly available datasets from simulation libraries are utilized, and its algorithm is designed to be dataset-agnostic and robust. However, it is essential to continually monitor the development of experimental datasets for 2D materials²⁶, as they are crucial for the future testing and fine-tuning of the DD2D, enabling it to adapt to a broader range of practical environments.

We hope this model can be applied in high-throughput analysis or screening systems to provide candidate structures in real-time, significantly accelerating the exploration of 2D materials. Furthermore, although the DD2D focuses on 2D structural reconstruction, such a physically sensitive model could pave the way for reverse design of deep learning models across broader materials fields.

Methods

Dataset preparation

The datasets used in this study were obtained from the C2DB database¹⁶, and thermodynamic perturbations for Ab Initio Molecular Dynamics (AIMD) were performed using the Vienna Ab-initio Simulation Package (VASP)^{27,28}. With a total of 7951 samples for training, the dataset was

divided into training, validation, and test subsets in a ratio of 8:1:1, resulting in 796 samples in the test set. During testing, the model recognizes the diffraction patterns and the corresponding crystal structures. It is noted that the original and perturbed datasets were independently trained.

A beam incident upon a structure containing N atoms is considered, where the coherent scattering intensity I_{hk} is defined by Eq. (1):

$$I_{hk} = I_0 \frac{AP}{\sin 2\theta} |AFF(q_{hk})|^2 S(q_{hk}) \quad (1)$$

Where the A is Debye-Waller unction, the P is Lorentz polarization function, and the $AFF(q_{hk})$ is atomic form factor²⁹. The structure factor $S(q_{hk})$ is expressed as follows:

$$S(q_{hk}) = \frac{1}{N} \sum_{j=1}^N \sum_{i=1}^N e^{-iq_{hk}(\mathbf{R}_i - \mathbf{R}_j)} \quad (2)$$

where \mathbf{R}_j is the atomic fraction coordinate of the j -th atom in the structure, and q is the scattering vector. In the realistic calculation, we only consider the results of the coherent signal to draw the $S(q_{hk})$ pattern¹¹.

In geometric configuration, the c -axis is oriented perpendicular to the plane of the 2D crystal. Thus, the lattice parameters (a, b, c) and their corresponding angles (α, β, γ) are specified as the initial six parameters. The atomic positions are arranged by atomic number, followed by sorting the Cartesian coordinates by the L2 norm. After fully populating all coordinate information, the corresponding elemental data is embedded at the end. The sequence encoding concludes with the arrangement of elemental information corresponding to each site. Comprehensive information about

the 2D materials is obtained by combining this with lattice parameters and site information. Diffraction spectra generally contain numerous point data, with their distribution and intensity indicating the atomic arrangement of materials. To preserve the local texture information, a point cloud data structure is employed to retain the characteristics of diffraction spectra as much as possible.

Training details

The DD2D contains two transformer-based deep learning encoders^{30–33}, and the CGE section receives a segment encoded crystal as input. Corresponding [CLS_star, CLS_end] tokens are strategically placed at the beginning and end of the sequence to supervise the representation learning of the lattice and atomic components³¹. After processing through the transformer layers, all information is concatenated and subjected to normalization. In the STE section, specific sites and intensity values from the point cloud data are separated and concatenated, and then the intensity values are logarithmically transformed. In the dual-tower framework, the sequence data received at both ends is fed into the transformer layers, where self-attention calculations are performed via Eq. (3), as:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

where the query (Q), key (K) and value (V) are obtained by performing a fully connected layer transformation on the input, with d_k being the dimension of embedding. After encoding through the dual-tower model³⁴, for a given sample pair of structure vs. diffraction pattern, we obtain two normalized embeddings: the Crystallographic Geometric Representation (G_i) and the Site Texture Representation (T_i)³². Given a batch of N structure vs. diffraction pattern sample pairs, our model is trained to predict which of the $N \times N$ possible structure vs. diffraction pattern sample pairings across a batch actually occurred. The objective of the model is to maximize the cosine similarity between the G_i and the T_i of the N real pairs in the batch, while minimizing the cosine similarity of the embeddings of the $N^2 - N$ incorrect pairings³¹. Thus, we optimize a symmetric cross entropy loss \mathcal{L} over these similarity scores, formulated as follows Eq. (4):

$$\mathcal{L} = -\frac{1}{2N} \sum_{i=1}^N \log\left(\frac{\exp(T_i \cdot G_i)}{\sum_{k \in [1, N]} \exp(T_i \cdot G_k)}\right) - \frac{1}{2N} \sum_{i=1}^N \log\left(\frac{\exp(T_i \cdot G_i)}{\sum_{k \in [1, N]} \exp(T_k \cdot G_i)}\right) \quad (4)$$

Furthermore, the ablation experiment examined the impact of key parameters in DD2D's Transformer backbone, including the number of attention heads and the number of Transformer layers. The detailed test results of different prediction attempt range (Attempts 1, 3, 5 and 10) were listed in Supplementary Table 1.

Data availability

Source data is archived under Zenodo³⁵. The dataset used to train the deep learning model is available at a GitHub repository [https://github.com/SuthPhy2Ai/DD2D] and archived under Zenodo³⁶.

Code availability

The complete code of this study is openly accessible via GitHub repository [https://github.com/SuthPhy2Ai/DD2D] along with a description, which includes instructions on how to run the code, required dependencies, and explanations of the main modules.

Received: 3 November 2024; Accepted: 20 May 2025;

Published online: 28 May 2025

References

- Novoselov, K. S. et al. Electric Field Effect in Atomically Thin Carbon Films. *Science*. **306**, 666–669 (2004).
- Liu, X. & Hersam, M. C. 2D materials for quantum information science. *Nat. Rev. Mater.* **4**, 669–684 (2019).
- Novoselov, K. S. et al. Two-dimensional gas of massless Dirac fermions in graphene. *Nature* **438**, 197–200 (2005).
- Ho, P. Twenty years of 2D materials. *Nat. Phys.* **20**, 1 (2024).
- Zhang, Y., Tan, Y.-W., Stormer, H. L. & Kim, P. Experimental observation of the quantum Hall effect and Berry's phase in graphene. *Nature* **438**, 201–204 (2005).
- Novoselov, K. S. et al. Two-dimensional atomic crystals. *Proc. Natl Acad. Sci.* **102**, 10451–10453 (2005).
- Goodwin, A. L. Opportunities and challenges in understanding complex functional materials. *Nat. Commun.* **10**, 4461 (2019).
- Gautam, S., Singhal, P., Chakraverty, S. & Goyal, N. Synthesis and characterization strategies of two-dimensional (2D) materials for quantum technologies: A comprehensive review. *Mater. Sci. Semicond. Process* **181**, 108639 (2024).
- Garrido, M., Naranjo, A. & Pérez, E. M. Characterization of emerging 2D materials after chemical functionalization. *Chem. Sci.* **15**, 3428–3445 (2024).
- Latychevskaia, T. Three-Dimensional Structure from Single Two-Dimensional Diffraction Intensity Measurement. *Phys. Rev. Lett.* **127**, 063601 (2021).
- Lucking, M. C. et al. Traditional Semiconductors in the Two-Dimensional Limit. *Phys. Rev. Lett.* **120**, 86101 (2018).
- Mujib, S. et al. Design, characterization, and application of elemental 2D materials for electrochemical energy storage, sensing, and catalysis. *Mater. Adv.* **1**, 2562–2591 (2020).
- Xu, R. et al. Advanced atomic force microscopies and their applications in two-dimensional materials: a review. *Mater. Futures* **1**, 032302 (2022).
- Keen, D. A. & McGreevy, R. L. Structural modelling of glasses using reverse Monte Carlo simulation. *Nature* **344**, 423–425 (1990).
- Albinati, A. & Willis, B. T. M. The Rietveld method. In *International Tables for Crystallography, Online MRW 710–712* (John Wiley & Sons, Ltd, 2006). <https://doi.org/10.1107/97809553602060000614>.
- Gjerding, M. N. et al. Recent progress of the Computational 2D Materials Database (C2DB). *2d Mater.* **8**, 044002 (2021).
- Lei, Y. et al. Graphene and Beyond: Recent Advances in Two-Dimensional Materials Synthesis, Properties, and Devices. *ACS Nanosci. Au* **2**, 450–485 (2022).
- Xie, T. & Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
- Reiser, P. et al. Graph neural networks for materials science and chemistry. *Commun. Mater.* **3**, 93 (2022).
- Yao, Z. et al. Machine learning for a sustainable energy future. *Nat. Rev. Mater.* **8**, 202–215 (2023).
- Chen, C. & Ong, S. P. A universal graph deep learning interatomic potential for the periodic table. *Nat. Comput. Sci.* **2**, 718–728 (2022).
- Hart, G. L. W., Mueller, T., Toher, C. & Curtarolo, S. Machine learning for alloys. *Nat. Rev. Mater.* **6**, 730–755 (2021).
- Zhang, S. et al. Crystallographic phase identifier of a convolutional self-attention neural network (CPICANN) on powder diffraction patterns. *IUCrJ* **11**, 634–642 (2024).
- Knøsgaard, N. R. & Thygesen, K. S. Representing individual electronic states for machine learning GW band structures of 2D materials. *Nat. Commun.* **13**, 468 (2022).
- Kaufman, L. & Rousseeuw, P. J. Finding groups in data: an introduction to cluster analysis. *J. Appl. Stat.* **51**, 1618–1620 (2005).
- Yang, J.-H. et al. Experimental data platform for 2D materials from synthesis to physical properties. *Digital Discov.* **3**, 573–585, <https://2DMat.ChemDX.org> (2024).

27. Kresse, G. & Hafner, J. Ab initio molecular dynamics for open-shell transition metals. *Phys. Rev. B* **48**, 13115–13118 (1993).
28. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
29. Brown, P. J., Fox, A. G., Maslen, E. N., O'Keefe, M. A. & Willis, B. T. M. Intensity of diffracted intensities. In *International Tables for Crystallography Volume C: Mathematical, physical and chemical tables* (ed. Prince, E.) 554–595 (Springer Netherlands, Dordrecht, 2004). <https://doi.org/10.1107/97809553602060000600>.
30. Vaswani, A. et al. Attention is All you Need. In *Advances in Neural Information Processing Systems* (eds Guyon, I. et al.) vol. 30 (Curran Associates, Inc., 2017).
31. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *North American Chapter of the Association for Computational Linguistics* (2019).
32. Radford, A. et al. Learning Transferable Visual Models From Natural Language Supervision. In *Proc. 38th International Conference on Machine Learning* (eds Meila, M. & Zhang, T.) vol. 139 8748–8763 (PMLR, 2021).
33. He, K., Fan, H., Wu, Y., Xie, S. & Girshick, R. Momentum Contrast for Unsupervised Visual Representation Learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* 9726–9735. <https://doi.org/10.1109/CVPR42600.2020.00975> (2020).
34. Huang, P.-S. et al. Learning deep structured semantic models for web search using clickthrough data. In *Proc. 22nd ACM International Conference on Information & Knowledge Management* (2013).
35. <https://doi.org/10.5281/zenodo.15289770>.
36. <https://doi.org/10.5281/zenodo.13921167>.

Acknowledgements

This work was supported financially by the National Natural Science Foundation of China (Grant No. 51690160), the Young Scientists Fund of the National Natural Science Foundation of China (Grant No. 12404276), the Special Funds of the National Natural Science Foundation of China (Grant No. 12347164), the China Postdoctoral Science Foundation (Grant No. 2024T170541), China Postdoctoral Science Foundation (Grant No. GZC20231535), AINSE Postgraduate Research Award (PGRA) and China Scholarship Council (CSC, No.201906890056).

Author contributions

M.L. and Z.R. conceived and designed the study. T.S., M.L. and R.F. performed the dataset and trained the model. R.F. and T.S. draw the pictures. R.F., T.S., M.L., Y.W., R.O., D.S., M.C., T.Z., S.H. and Z.R. analyzed

the results of data. S.H. performed the ablation experiments. R.F., T.S., M.L., T.Z. and Z.R. wrote the manuscript. All authors discussed the results and contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42005-025-02152-8>.

Correspondence and requests for materials should be addressed to Musen Li or Zhongming Ren.

Peer review information *Communications Physics* thanks Pawan Tripathi, Aleksandr Zaloga and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025