npj Computational Materials

www.nature.com/npjcompumats

PERSPECTIVE OPEN

Check for updates

Off-the-shelf deep learning is not enough, and requires parsimony, Bayesianity, and causality

Rama K. Vasudevan[1]✉, Maxim Ziatdinov[2], Lukas Vlcek[3,4] and Sergei V. Kalinin[1]✉

Deep neural networks ('deep learning') have emerged as a technology of choice to tackle problems in speech recognition, computer vision, finance, etc. However, adoption of deep learning in physical domains brings substantial challenges stemming from the correlative nature of deep learning methods compared to the causal, hypothesis driven nature of modern science. We argue that the broad adoption of Bayesian methods incorporating prior knowledge, development of solutions with incorporated physical constraints and parsimonious structural descriptors and generative models, and ultimately adoption of causal models, offers a path forward for fundamental and applied research.

npj Computational Materials (2021)7:16 ; https://doi.org/10.1038/s41524-020-00487-0

INTRODUCTION

The spectacular growth of deep learning (DL) in the last decade has fueled the rise of a wave of data science and artificial intelligence ('AI') that has already had global impact across society. The spectacular successes of deep learning in some traditionally very difficult tasks in computer vision, natural language processing, machine translation, speech recognition, and gameplay has piqued interest across all scientific communities[1]. Here, deep learning refers to an approach that utilizes artificial neural networks (which have been available for decades[2] and have been intermittently used for specific problems[3,4]) that are comprised of numerous layers of stacked artificial neurons, and with oftentimes millions of trainable parameters, usually to approximate some highly complex nonlinear function. The networks are usually trained using a backpropagation algorithm and stochastic gradient descent to adjust the weights of the network to minimize some objective function, and nowadays expressly run on graphical or tensor processing units optimized for such calculations.

Depending on the specific architectures involved, deep neural networks (DNN) can be used in tasks including classification, regression (for instance, material property predictions based on a material's structure), as well as for unearthing correlations and compressing data in large datasets. A simple example of a DNN used in a materials science setting is shown in Fig. 1: in this case this network has been 'trained' to automatically identify atoms from noisy electron microscopy images. The network was trained by ingesting large volumes of simulated electron microscopy images where the atomic positions are known and therefore used as the 'labeled' data. In this process the network's parameters are continually updated to minimize the discrepancy between the predictions of the network and the ground truth (the positions of the atoms). The network can then be fed an image that was not part of the original training set to give the output of the atomic coordinates present, thereby operating as an automatic atom finder. In addition to simple image segmentation tasks, DNNs have also seen success in the trickier task of "generative modeling," which refers to the ability to generate datapoints (samples) that are not in the original dataset[5].

The key distinction between traditional machine learning (ML) and modern deep learning is that deep neural networks learn representations ('features') of the data as part of the training process, as opposed to being hand-crafted by domain experts, which was the prevailing method prior to the DL revolution. However, this also presents a problem: are the representations learned by the existing DL methods useful for aiding in understanding of physics and materials science? Even from a computer science perspective, DL, for all its successes, is surprisingly fragile and highly susceptible to adversarial attacks[6], in which input data are slightly perturbed in subtle ways that slowly guide the network to mis-classify the data with near 100% certainty[7]. A recent example shows that a DL-trained classifier of objects can mis-classify simple objects merely if they are displayed in specific unseen poses[8]. How can we then 'trust' the predictions of DL-based models, when they appear highly fragile and vulnerable? Perhaps as a less exotic example, how do we know which network architecture will give a correct, quantitative answer for a specific problem, and how can we quantify uncertainties and systematic and random errors in such an answer?

WHEN DOES ML WORK?

From the early days of machine learning, it was repeatedly noted that ultimately ML and DL serve as universal interpolators, finding correlations between large datasets in multidimensional spaces. At the same time, the physical sciences are based on the notion of hypothesis-driven science, often using observations from a set of experiments to reveal correlations, explore causal relationships, and ultimately unveil the underpinning physical laws. Thus, when and how can ML and AI methods be used to explore physics?

We note that the pitfalls of the conventional correlative modeling and their consequences are well explored[9,10]. The classical examples include Simpson paradox[11], where for example it is possible that statistically, a certain drug can be beneficial for humans in general, but detrimental for both males and females. While in areas such as sociology, medicine, and economics the approaches to deal with these issues are well developed, this is generally not the case in the physical sciences. Notably, the use of

[1]Center for Nanophase Materials Sciences, Oak Ridge, TN 37831, USA. [2]Computational Sciences and Engineering Division, Oak Ridge, TN 37831, USA. [3]Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [4]Present address: Bayer, St. Louis, MO 63141, USA. ✉email: vasudevanrk@ornl.gov ; sergei2@ornl.gov

Published in partnership with the Shanghai Institute of Ceramics of the Chinese Academy of Sciences

npj nature partner journals

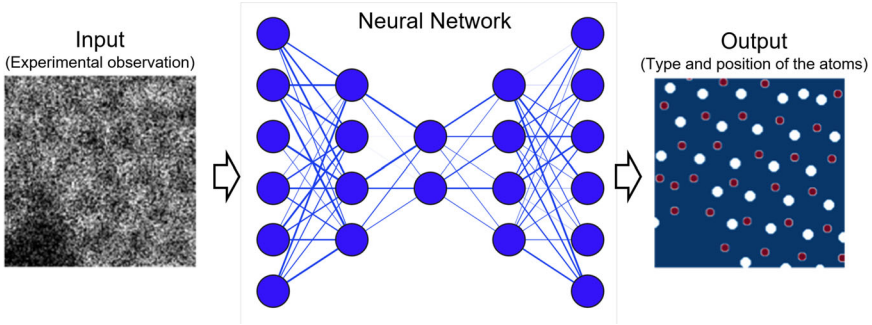

Fig. 1 Deep neural network for analyzing atomically resolved data by performing a semantic image segmentation[49,50] on the atomic level. The network accomplishes two goals: (i) it removes noise and (ii) it separates different atomic species into different classes on the level of individual image pixels. The input image is the scanning transmission electron microscopy image of disordered atomic lattice of 2D boron nitride; the output is the atomic coordinates.

complex machine learning models will not compensate for the incorrect causative attribution and would rather make the problem less obvious and more difficult to identify.

Here, we argue that the causal framework developed by Judea Pearl and expanded by Scholkopf, Mooij, and others provides a clear pathway towards answering these questions[9,12,13]. Generally, ML methods provide a universal and extremely powerful framework for analysis of physical problems when the causal chain is clearly known. The use of neural networks for the analysis of atomically resolved images[14] is causally determined, since the point-like objects observed in electron microscope at this level of resolution can only be atoms. In comparison, these models will not generalize for all images, since large collection of sphere-like objects can also describe chain mail, cloth, meshes, structure of certain minerals, etc. Similarly, the use of generative adversarial networks for the analysis of the simulated 4D scanning transmission electron microscopy data (STEM), or classical back-propagation networks to identify Ising model parameters based on hysteresis loops[15,16] is causally determined, since there is clear causal relationship between the inputs and outputs. At the same time, such trained networks can fail when applied to experimental data, since the instrument parameters are a biasing factor. In some cases, these can be accounted for by scaling and normalization, but not so in others, where calibration factors are numerous and the effect on the image is much more complex, and hence need to be calibrated in advance. Parenthetically, the outstanding success of DL learning methods as applied in the theoretical domain owes to the fact that the causal links there are explicit.

At the same time, ML methods can be expected to fail, and often fail, in cases where the causal links are uncertain. This includes multiple variants, including the presence of confounding factors that affect both (input) *X* and (output) *Y*, observational biases, etc. (see Fig. 2). Correspondingly in these cases the ML model, no matter how good, will fail to predict and generalize since there are control factors outside of the model. For instance, if a material property is predicted by ML models on the basis of only local structure and global chemistry (and not local chemical environments), this can easily lead to erroneous predictions in cases where it is the local chemical environment driving the changes in the first place. Then, the question is, does machine learning here become useless? Interestingly, the answer is that is still extremely useful – as long as the model is used in the parameter space in which the confounding factors are constant and observations are made with the same biases.

So, what are the other areas for ML in physics, beyond the conditioned correlative models valid when the causal links are known or defined? One class of the models is those that explore the complexity of the dataset, either via manifold learning in purely data spaces, or symbolic reconstructions, or extraction of generative models. These models exploit the fact that physical

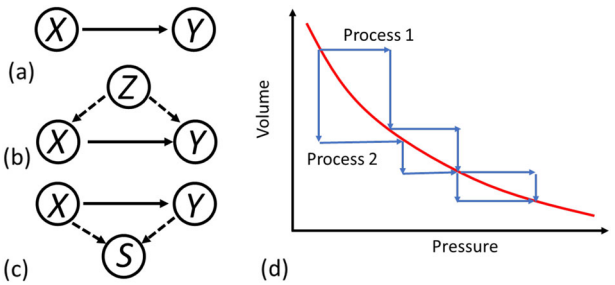

Fig. 2 Understanding causal links in the presence of confounding effects and observational biases. a Correlation can be used to analyze causative mechanisms only if there is a well-defined causal link between the variables. In the presence of (**b**) confounders or (**c**) observational bias analysis of correlations can result in fundamentally incorrect conclusions. For example, the correlation of the level of chocolate consumption and Nobel prize winning does not imply that chocolate can be used to increase scientific visibility; rather the same factors that enable higher consumption also lead to a higher probability of winning[51]. **d** Example of exploring causative mechanisms in physics. Observations will generate the correlation between pressure and volume. The analysis of the functional relationship between the two will yield the ideal gas law. With that, note that the knowledge of functional relationship is insufficient to analyze the causal mechanisms: does the pressure change cause changes in volume, or vice-versa?

laws are generally parsimonious. As an example, consider the use of neural networks with constraints placed on learned representations to answer a scientific hypothesis – that of a heliocentric solar system[17]. As analyzed by Lin, Tegmark and Rolnick[18], the success of deep learning is inherently linked to the fact that most complex systems, including those in physics, are hierarchical and are drawn from a very small subset of all possible data distributions.

Learning meaningful representations: looking for simplicity

DL methods learn a representation of the inputs that is advantageous to the task that is required to be performed, which are sometimes referred to as 'features.' Are features learned by such networks physically reasonable or at all meaningful for materials scientists? After all, the predictions of a DNN may be highly accurate, but might have little to no extrapolation ability. This is because the features learned are the basis used for predictions of the model, and physically non-meaningful features can lead to highly inaccurate predictions for unseen data.

We argue that one method to aid the learning of *better* representations of systems is to incorporate principles from statistical physics. To be truly predictive, and not just interpolative, DNNs need to carry an internal representation of the physical

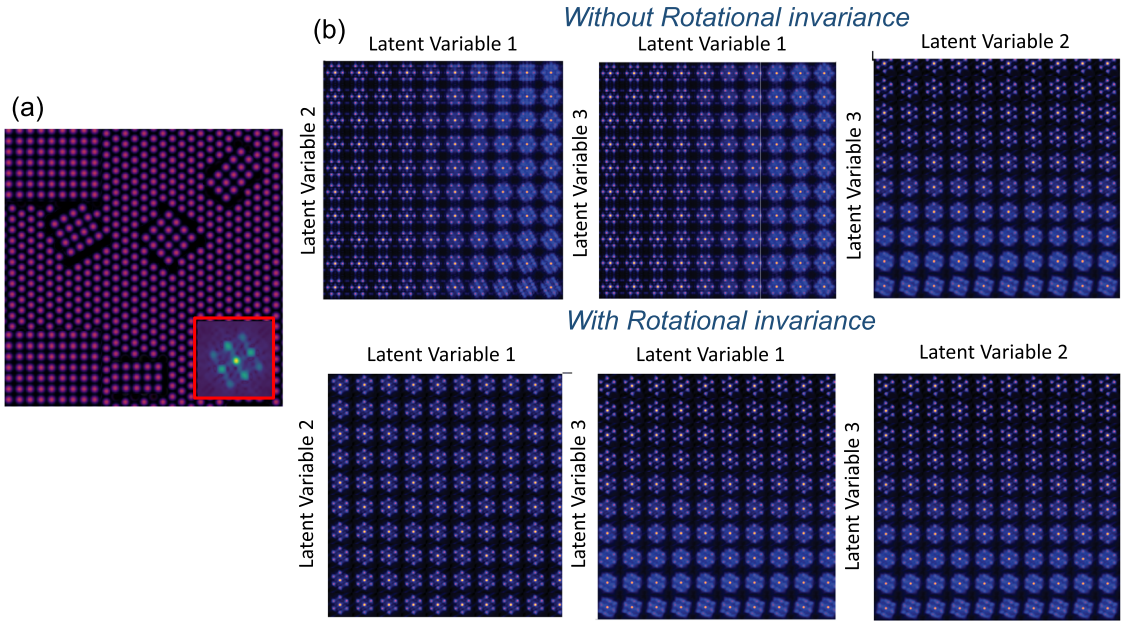

Fig. 3 Utility of adding rotational invariances when forming reduced representations. a A simulated test image (size 500 × 500 px) containing different lattice structures with arbitrary rotations. We aim to disentangle rotations of a structure from the actual structures themselves. For this purpose, we begin by computing a sliding window 2D Fourier transform[52] over the image, collecting hundreds of 2D local Fourier transformed images. We then aim to encode this dataset using a variational autoencoder. Shown in (**b**), (**c**) are the latent space visualizations for the encoder (**b**) without rotational invariance, and (**c**) with rotational invariance specifically built-in to the loss function. It is easily observed that without rotational invariance, the latent space is forced to take into account rotations, decreasing the compactness of the representation. The latent space in (**c**), however, is much simpler (parsimony).

system, which is ultimately given by its microstate probability distribution or partition function. Measured properties are then derived as specific projections (i.e., coarse-graining) of the distribution. For example, for a 2D Ising model, the property of note (magnetization) can be derived in this way[19]. The means to implement this can vary; however, the core aim is to ensure that the predictions are physically meaningful in terms of microstate probabilities.

Moreover, one can employ regularization that is relevant for physics, in that we ensure such physical representations are, and must be expressed as only a small number of independent latent variables, that is, be parsimonious[17]. For example, consider the case of rotational symmetry. It is possible to incorporate this feature directly into neural networks, such as variational autoencoders[20], such that multiple rotational variants are all mapped to the same latent space descriptors learned by the network, as shown in Fig. 3 (image and associated code is available in supplementary information). Here, the "building blocks" (latent space vector) is sampled and plotted (Fig. 3b, c), essentially highlighting what was 'learned' by the network. In the case of the rotationally invariant variational autoencoder, multiple rotational variants are all mapped to the same rotation, and the latent space learns varies in terms of intensity and type of diffraction pattern, whereas for the rotationally variant case rotations must be encoded specifically in the latent space, leading to a more complicated representation. Beyond incorporation of symmetry, the actual type of loss function itself can be augmented as a form of physics-based regularization (Fig. 4). For instance, instead of regressing on mean squared error of inputs and outputs directly, the loss can be computed in a space that more directly captures the behavior of a thermodynamic system, i.e., the configuration space. Instead of a mean squared error, the statistical distance metric can be employed, which is related to distinguishability of thermodynamic systems. Of course, the challenge here is in determining the specific features, but again, this may be learned (for example using generative adversarial networks or autoencoders). We note as an aside here that the links between neural networks and statistical physics, and the field of statistics more generally, go back at least three decades[21].

Adding context: Bayesian methods and prior knowledge

Another major class of models are the Bayesian models. While DL requires large volumes of data and attempts to learn representations without the need for priors (beyond those encoded within the architecture design, such as convolutions which introduce spatial invariance), this is not the case for most physical problems. Indeed, the question of most importance is how best to incorporate prior knowledge of scientists within a data-driven approach.

The natural approach for incorporation of the past knowledge in the analysis is based on Bayesian methods, derived from the celebrated Bayes formula:

$$p(\theta_i|D) = \frac{p(D|\theta_i)p(\theta_i)}{p(D)} \quad (1)$$

Here D represents the observed data, $p(D|\theta_i)$ is the likelihood that the data can be generated by the theory, i.e., the model, i, with parameters, θ. The prior knowledge is represented by $p(\theta_i)$. Finally, $p(D)$ is the denominator that defined the total space of possible outcomes. Despite the elegance and transparency of Bayesian approach, its adoption by many scientific communities has been rather slow. First, evaluation of denominator in Eq. 1 requires very high dimensional integrals and become feasible for experimentally relevant distributions only over the last decade. Secondly, the choice of the priors represents an obvious issue. Interestingly, in the physics field, domain knowledge is typically abundant, necessitating translating of past domain knowledge into the language of probability distribution functions. In a sense, Bayes formula represents the synergy of experimental science as a source of data, domain expertise as source of priors, theory as a source of likelihoods, and high-performance computing necessary to address the associated computational challenges.

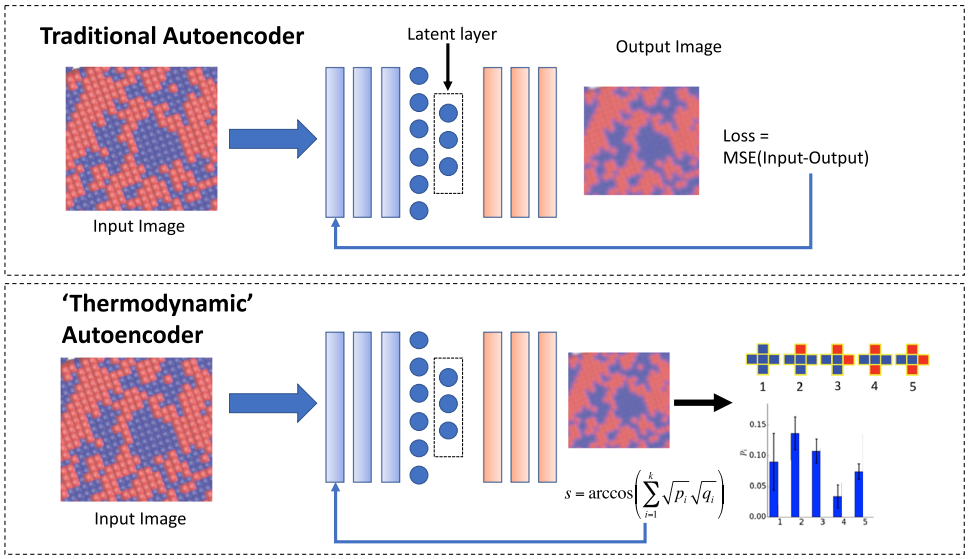

Fig. 4 Adjusting loss functions to emphasize physically meaningful comparisons. A traditional autoencoder contains a loss function that minimizes the mean-squared error between the input and output image or image batches. However, in learning representations for thermodynamic systems, it may be preferable not to output the same image – but rather, to capture the essence of the image, i.e., the statistical fluctuations of the configurations of species present. In such cases, the loss function can be appropriately modified, such as with a statistical distance loss function that computes the distance between two systems based on the probability of observing different configurations. Of course, the challenge here is that the configurations to compare should themselves be learned. That may be possible using dual-network architectures, such as generative adversarial networks where one component determines the features that maximize distinguishing of the two systems, while the other network aims to minimize the discrepancy.

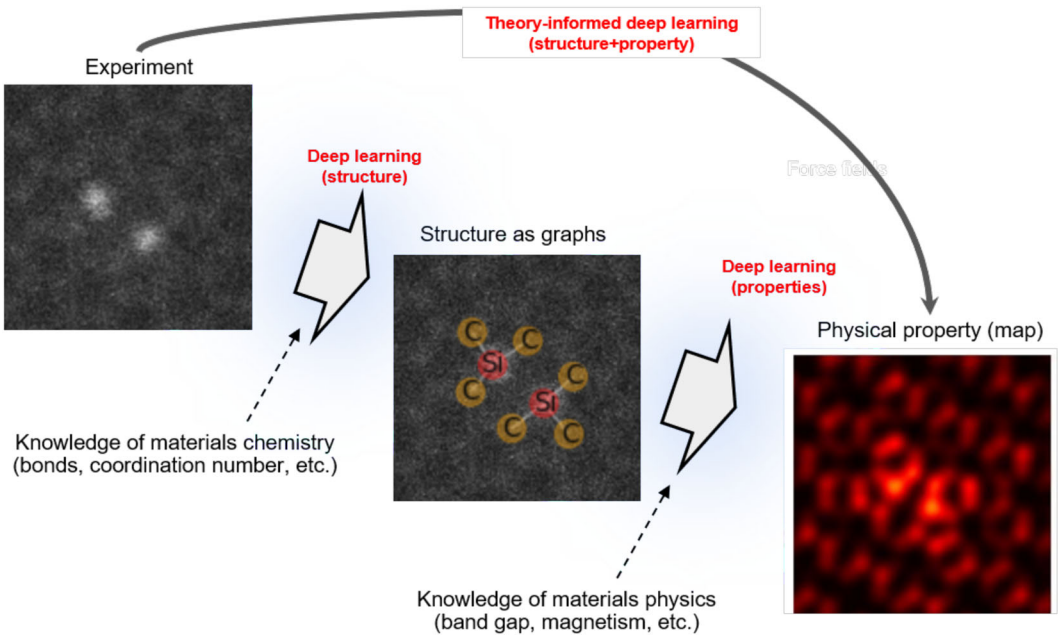

Fig. 5 A simplified schematic showing analysis of scientific image data using a combination of deep learning and graph modeling for predicting materials (local) structure and properties. Knowledge can be injected at both the structural learning step, as well as the translation from structure to physical properties.

The adoption of Bayesian approaches allows us to systematically explore complex problems, fusing prior information from other sources. For example, in a scientific image processing task, can the neural network performance be improved based on knowledge on which functional groups are possible for specific materials class, and their relevant energies/probabilities? The combination of a convolutional neural networks with graphical models[22] (e.g., Markov random field) may allow incorporating prior knowledge about physiochemical properties of a system, such as a probability of realization of certain lattice-impurity configurations, into the decoding of experimental observations (see Fig. 5). Deep learning models such as graph convolutional neural networks now also allow predicting materials properties directly from crystal lattice graphs[23]. However, this approach is

currently limited to ideal periodic systems. Predicting the local property maps (e.g., distribution of local density of states) directly from the experimental observations is a major challenge.

Other methods to leverage or couple physical models and machine learning have been proposed. For instance, the 'theory-driven data science' paradigm espoused by Karpatne et al.[24] describe several such approaches, including pre-training of networks with simulated data from physical models (so that when trained on real-world data, the networks are more likely to yield physically plausible results and can be used for deconvolution of experimental parameters[15,16]), and constrained optimization, where solutions must obey constraints such as being valid solutions to a partial differential equation. Determining the most effective methods to encode these relationships within deep neural networks remains an ongoing challenge, and invariably a tension between the flexibility of the model and the ability to learn physically meaningful relationships that underpin extrapolation ability will exist.

Of course, in some instances it may be better to avoid prior information entirely: for example, the AlphaZero[25] program mastered the games of Chess, Shogi, and Go starting from random play and given no domain knowledge other than the game rules, and yet achieved super-human performance in all three. This can allow for strategies to emerge that humans may not have envisioned[26]. We foresee the utility of these approaches in particular to areas such as controlled materials synthesis, drug discovery[27], and other design spaces[28].

Data and DL future

Finally, we explore the changes in scientific community and infrastructure needed to make this deep learning transformation possible. Most of the critical algorithmic developments for deep learning, such as convolutional networks and back propagation, occurred decades ago[29]. Rather, it was the availability of large, labeled databases, and the ability to compute these huge volumes to enable network training, that were key factors in the current deep learning revolution[30]. As such, the development of open source libraries of materials data is an instrumental part, and a slew of recent reviews[31–33] touch on the need and benefits of these databases.

The adoption of machine learning tools, including basic knowledge and relevant programming skills by the broad scientific community is becoming a necessity. A related issue is the availability and distribution of tested, well-documented codes. While GitHub and Jupyter notebooks[34] offer an effective means for code sharing, development, and universal access, the incentive system in fundamental science is heavily tilted towards publication as a primary measure of performance. Correspondingly, increasing the visibility of code development and re-use, and ideally integrating codes into scientific publications (e.g., see ref. [35]) becomes more important. Ultimately, data, code, and workflow sharing will become the primary pathway for collaboration and scientific knowledge dissemination, complementing, and potentially surpassing archival publications.

Overall, the initial forays in machine learning across physical science communities have demonstrated the power of these methods in a variety of domains. But practical implementation will require additional work on adjusting the tools to match the problems presented in those areas. In our opinion, the integration of human domain expertise and causal inference with deep learning will be the crucial link to correctly harnessing and exploiting the benefits that DL and ML can provide. Most importantly, the merger of machine learning with classical hypothesis driven science can bring ML beyond the current correlative paradigms into larger fields of Bayesian and causal learning and establish connections to the materials world via automated experiment[36–48] and open instrumental facilities, thus giving rise to fundamentally different ways of scientific research.

DATA AVAILABILITY

The image in Fig. 3 is provided in the supplementary information.

CODE AVAILABILITY

The script to analyze the image in Fig. 3 is provided in the supplementary information.

Received: 1 May 2020; Accepted: 3 December 2020;
Published online: 27 January 2021

REFERENCES

1. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
2. Rosenblatt, F. *Principles of Neurodynamics. Perceptrons and the Theory of Brain Mechanisms* (Cornell Aeronautical Lab Inc Buffalo NY, 1961).
3. Lee, K. K. et al. Using neural networks to construct models of the molecular beam epitaxy process. *IEEE Trans. Semicond. Manuf.* **13**, 34–45 (2000).
4. Sumpter, B. G. & Noid, D. W. Potential energy surfaces for macromolecules. a neural network technique. *Chem. Phys. Lett.* **192**, 455–462 (1992).
5. Doersch, C. Tutorial on variational autoencoders. Preprint at https://arxiv.org/abs/1606.05908 (2016).
6. Su, J., Vargas, D. V. & Sakurai, K. One pixel attack for fooling deep neural networks. *IEEE Trans. Evol. Comp.* **23**, 828–841 (2019).
7. Warde-Farley, D. & Goodfellow, I. In *Perturbations, Optimization, and Statistics* (eds Hazan, T., Papandreou, G. & Tarlow, D.) Ch. 11 (MIT Press, 2016).
8. Alcorn, M. A. et al. Strike (with) a pose: neural networks are easily fooled by strange poses of familiar objects. *Proc. IEEE Conf. Comp. Vis. Patt. Recog.* 4845–4854 (IEEE, New York, 2019).
9. Pearl, J. *Causality* (Cambridge Univ. Press, 2009).
10. O'neil, C. *Weapons of Math Destruction: How big data Increases Inequality and Threatens Democracy* (Broadway Books, 2016).
11. Wagner, C. H. Simpson's paradox in real life. *Am. Statistician* **36**, 46–48 (1982).
12. Hoyer, P. O., Janzing, D., Mooij, J. M., Peters, J. & Schölkopf, B. Nonlinear causal discovery with additive noise models. *Adv. Neur. Inf. Proc. Sys.* **21**, 689–696 (2009).
13. Peters, J., Janzing, D. & Schölkopf, B. *Elements of Causal Inference: Foundations and Learning Algorithms* (MIT Press, 2017).
14. Ziatdinov, M. et al. Deep learning of atomically resolved scanning transmission electron microscopy images: chemical identification and tracking local transformations. *ACS Nano* **11**, 12742 (2017).
15. Ovchinnikov, O., Jesse, S., Bintacchit, P., Trolier-McKinstry, S. & Kalinin, S. V. Disorder identification in hysteresis data: recognition analysis of the random-bond–random-field ising model. *Phys. Rev. Lett.* **103**, 157203 (2009).
16. Kumar, A. et al. Spatially resolved mapping of disorder type and distribution in random systems using artificial neural network recognition. *Phys. Rev. B.* **84**, 024203 (2011).
17. Iten, R., Metger, T., Wilming, H., Del Rio, L. & Renner, R. Discovering physical concepts with neural networks. *Phys. Rev. Lett.* **124**, 010508 (2020).
18. Vlcek, L., Maksov, A. B., Pan, M., Vasudevan, R. K. & Kalinin, S. V. Knowledge extraction from atomically resolved images. *ACS Nano* **11**, 10313–10320 (2017).
19. Koch-Janusz, M. & Ringel, Z. J. N. P. Mutual information, neural networks and the renormalization group.*Nat. Phys.* **14**, 578–582 (2018).
20. Kingma, D. P. & Welling, M. Auto-encoding variational bayes. Preprint at https://arxiv.org/abs/1312.6114 (2013).
21. Peretto, P. Collective properties of neural networks: a statistical physics approach. *Biol. Cybern.* **50**, 51–62 (1984).
22. Arnab, A. et al. Conditional random fields meet deep neural networks for semantic segmentation: combining probabilistic graphical models with deep learning for structured prediction. *IEEE Signal Process. Mag.* **35**, 37–52 (2018).
23. 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).
24. Karpatne, A. et al. Theory-guided data science: a new paradigm for scientific discovery from data. *IEEE Trans. Knowl. Data Eng.* **29**, 2318–2331 (2017).
25. Silver, D. et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* **362**, 1140–1144 (2018).
26. Berner, C. et al. Dota 2 with large scale deep reinforcement learning. Preprint at https://arxiv.org/abs/1912.06680 (2019).

27. Popova, M., Isayev, O. & Tropsha, A. Deep reinforcement learning for de novo drug design. *Sci. Adv.* **4**, eaap7885 (2018).
28. Zhou, Z., Kearnes, S., Li, L., Zare, R. N. & Riley, P. Optimization of molecules via deep reinforcement learning. *Sci. Rep.* **9**, 1–10 (2019).
29. Hecht-Nielsen, R. *Neural Networks for Perception* (Elsevier, 1992).
30. Gershgorn, D. *Quartz*. https://qz.com/1034972/the-data-that-changed-the-direction-of-ai-research-and-possibly-the-world/ (2017).
31. Mannodi-Kanakkithodi, A. et al. Scoping the polymer genome: a roadmap for rational polymer dielectrics design and beyond. *Mater. Tod.* **21**, 785–796 (2018).
32. Ramakrishna, S. et al. Materials informatics. *J. Intell. Manuf.* **30**, 2307–2326 (2019).
33. Rickman, J., Lookman, T. & Kalinin, S. Materials informatics: from the atomic-level to the continuum. *Acta Mater.* **168**, 473–510 (2019).
34. Perez, F. & Granger, B. E. *Project Jupyter: Computational Narratives as the Engine of Collaborative Data Science*. https://blog.jupyter.org/project-jupyter-computational-narratives-as-the-engine-of-collaborative-data-science-2b5fb94c3c58 (2015).
35. Ziatdinov, M., Nelson, C., Vasudevan, R. K., Chen, D. & Kalinin, S. V. Building ferroelectric from the bottom up: the machine learning analysis of the atomic-scale ferroelectric distortions. *Appl. Phys. Lett.* **115**, 052902 (2019).
36. Gómez-Bombarelli, R. et al. Automatic chemical design using a data-driven continuous representation of molecules. *ACS Cent. Sci.* **4**, 268–276 (2018).
37. Gómez-Bombarelli, R. et al. Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach. *Nat. Mater.* **15**, 1120–1127 (2016).
38. Sanchez-Lengeling, B. & Aspuru-Guzik, A. J. A. Inverse molecular design using machine learning: generative models for matter engineering. *Science* **361**, 360–365 (2018). c. s.
39. Tabor, D. P. et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nat. Rev. Mater.* **3**, 5–20 (2018).
40. Nikolaev, P. et al. Autonomy in materials research: a case study in carbon nanotube growth. *NPJ Comp. Mater.* **2**, 16031 (2016).
41. Noack, M. M., Doerk, G. S., Li, R., Fukuto, M. & Yager, K. G. Advances in kriging-based autonomous X-ray scattering experiments. *Sci. Rep.* **10**, 1–17 (2020).
42. Noack, M. M. et al. A kriging-based approach to autonomous experimentation with applications to X-ray scattering. *Sci. Rep.* **9**, 1–19 (2019).
43. Kusne, A. G. et al. On-the-fly machine-learning for high-throughput experiments: search for rare-earth-free permanent magnets. *Sci. Rep.* **4**, 1–7 (2014).
44. Gromski, P. S., Henson, A. B., Granda, J. M. & Cronin, L. How to explore chemical space using algorithms and automation. *Nat. Rev. Chem.* **3**, 119–128 (2019).
45. Steiner, S. et al. Organic synthesis in a modular robotic system driven by a chemical programming language. *Science* **363**, eaav2211 (2019).
46. Henson, A. B., Gromski, P. S. & Cronin, L. Designing algorithms to aid discovery by chemical robots. *ACS Cent. Sci.* **4**, 793–804 (2018).
47. Campbell, Z. S. & Abolhasani, M. Facile synthesis of anhydrous microparticles using plug-and-play microfluidic reactors. *React. Chem. Eng.* **5**, 1198–1211 (2020).
48. Epps, R. W., Felton, K. C., Coley, C. W. & Abolhasani, M. Automated microfluidic platform for systematic studies of colloidal perovskite nanocrystals: towards continuous nano-manufacturing. *Lab a Chip* **17**, 4040–4047 (2017).
49. Ronneberger, O., Fischer, P. and Brox, T. U-net: convolutional networks for bio-medical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234–241).
50. Ziatdinov, M. *AtomAI*. https://github.com/ziatdinovmax/atomai (2020).
51. Dablander, F. An introduction to causal inference. Preprint at https://psyarxiv.com/b3fkw/ (2020).
52. Vasudevan, R. K., Ziatdinov, M., Jesse, S. & Kalinin, S. V. Phases and interfaces from real space atomically resolved data: physics-based deep data image analysis. *Nano Lett.* **16**, 5574–5581 (2016).

ACKNOWLEDGEMENTS

The work was supported by the U.S. Department of Energy, Office of Science, Materials Sciences and Engineering Division (S.V.K., L.V., R. K. V.). Research was conducted at the Center for Nanophase Materials Sciences, which also provided support (M.Z.) and is a US DOE Office of Science User Facility.

AUTHOR CONTRIBUTIONS

R.K.V. wrote the outline and the majority of the perspective. S.V.K. provided the initial concept and participated in development of the ideas and assisted in manuscript preparation and developed the causality section. L.V. wrote sections regarding connections between statistical physics and machine learning. M.Z. wrote sections on contextual information and theory guidance.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information is available for this paper at https://doi.org/10.1038/s41524-020-00487-0.

Correspondence and requests for materials should be addressed to R.K.V. or S.V.K.

Reprints and permission 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.