Discovering invariants via machine learning

Seungwoong Ha and Hawoong Jeong ⊙*

Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 34141, Korea

Invariants and conservation laws convey critical information about the underlying dynamics of a system, yet it is generally infeasible to find them from large-scale data without any prior knowledge or human insight. We propose ConservNet to achieve this goal, a neural network that spontaneously discovers a conserved quantity from grouped data where the members of each group share invariants, similar to a general experimental setting where trajectories from different trials are observed. As a neural network trained with an intuitive loss function called noise-variance loss, ConservNet learns the hidden invariants in each group of multidimensional observables in a data-driven end-to-end manner. Our model successfully discovers underlying invariants from the simulated systems having invariants as well as a real-world double-pendulum trajectory. Since the model is robust to various noises and data conditions compared to the baseline, our approach is directly applicable to experimental data for discovering hidden conservation laws and further, general relationships between variables.

I. INTRODUCTION

Modern science greatly depends on the mathematical modeling of given systems and finding the internal structures between observables. One of the most important concepts in system modeling is the *invariants* that underlie the system dynamics, which provide significant information about structural symmetries and low-dimensional embeddings of the system. Invariants and symmetries are fundamental building blocks of nearly all physical systems in nature, such as classical systems with Hamiltonians, gauge orbits, and many other dynamical systems. Scientists have long attempted to identify the hidden correlations and interactions among the observables of such systems by discovering the conserved quantities and underlying symmetries.

Recently, with the advent of large-scale data and phenomenal advances in machine learning in physical sciences [1–12], various studies have contributed towards the *automation of science* [13], referring to current efforts to reveal scientific concepts and construct models solely from observed data without human intervention [14–24]. Following this line, several studies have attempted to accomplish the automated discovery of conserved quantities with neural networks [21–24]; limitations of these works though include the requirements for additional nonautomated preprocessing and often a great number of data sets from different conditions, as well as the ability to only infer the number of invariants. Real-

world empirical data are often sparse, noisy, and scattered into small groups, and hence a model for automated discovery needs to be robust to such harsh conditions.

In this study we introduce ConservNet, a neural network to discover conserved quantities in grouped data, such as trajectories, without any prior knowledge of the system. Instead of explicitly restricting the model to ensure certain symmetries, we propose a loss function that facilitates the model to directly learn the invariant function. We show that ConservNet robustly finds an invariant by reducing the intragroup variance of its output while preventing convergence into trivial constant functions. Our model can be applied to a variety of realistic data conditions with multiple groups, is robust to noises and nuisance variables, and employs a pipeline from raw data to invariants in an end-to-end manner that enables the direct extraction of symbolic formulas. We examine the capability of ConservNet by applying it to five model systems ranging from synthetic invariants to physical models that cover diverse functional forms, along with experimental trajectory data of a double pendulum. The robustness of our method clearly demonstrates the potential of ConservNet to be applied to real systems where data are sparse and no conservation laws are known.

II. NOISE-VARIANCE LOSS

Throughout this paper, the data condition (N, M) indicates that the data are divided into N groups, in which each group shares the same invariant and has M data points. Our goal is to find conserved quantities hidden in such grouped d-dimensional data that are expected to have at least one invariant. We assume that the system has an invariant function V that satisfies $V(\mathbf{x}_{ij}) = C_i$ for all $\mathbf{x}_{ij} \in G_i$, where G_i denotes the ith group and $\mathbf{x}_{ij} \in \mathbb{R}^d$ is the jth input data of dimension d from group i.

In order for the model to properly approximate the invariant, it needs to satisfy two important criteria. First, the desired model should produce a ground-truth invariant C, or at least a value strongly correlated with the true invariant. Second, the model output from the same group should be equal in the ideal case, or at least its deviation should be minimized.

To satisfy the second criterion, the loss function L for the neural output F_θ should decrease the intragroup variance of the outputs from each group, and thus the variance term $\mathcal{L}_{i,\text{var}} = M^{-1} \sum_j F_\theta(\mathbf{x}_{ij})^2 - [M^{-1} \sum_j F_\theta(\mathbf{x}_{ij})]^2$ should be minimized. Here the naive optimization of this loss function will generally fall into trivial minima. As an example, the whole class of simple multivariate functions $f : \mathbb{R}^d \to C_0$ for any real value C_0 becomes one of the global minima of $\mathcal{L}_{i,\text{var}}$ since the output is constant regardless of the input. Convergence to such a trivial solution would violate the first criterion in our case.

Thus, we need to guide F_θ to capture a nontrivial invariant besides a constant function. In this study we inhibit trivial convergence by adopting a *spreading* term that increases the variance of the output from improper input, such as perturbed input with noise. This spreading loss can be expressed as $\mathcal{L}_{i,\text{noise}} = |Q - \text{Var}[F_\theta(\mathbf{x}_{ij} + \boldsymbol{\varepsilon}_{ij})]|$, where Q is the spreading constant and $\boldsymbol{\varepsilon}_{ij}$ denotes a random noise vector, whose L_2-norm is bounded to $R = \max(\|\boldsymbol{\varepsilon}_{ij}\|_2)$. Here Q restricts the absolute value of the variance of the outputs from perturbed inputs, since optimization without this constraint will lead F_θ into a diverging function, ignoring the variance minimization term. Thus, the relative scale of Q and R controls the fineness of the spreading. Similar intuition for spreading loss can be found in a contrastive loss in self-supervised learning [25–27], which also needs to increase the distance in representation space between different classes while preventing divergence. Combining two terms and summing over all groups, the loss function for ConservNet becomes

$$\mathcal{L} = \sum_i \mathcal{L}_i = \sum_i \text{Var}[F_\theta(\mathbf{x}_{ij})] + |Q - \text{Var}[F_\theta(x_{ij} + \boldsymbol{\varepsilon}_{ij})]|. \tag{1}$$

We propose this loss function for capturing an invariant as noise-variance (NV) loss, as schematically depicted in Fig. 1. We prove that two competing terms in NV loss inhibits trivial convergence by preventing the gradient ∇F_θ from becoming $\vec{\mathbf{0}} \in \mathbb{R}^d$ [28], whose implication can be physically interpreted if the system has a well-defined Hamiltonian H. In the language of Hamiltonian mechanics, the model aims to learn a *constant of motion* G with various energy levels, which is a generating function of the (infinitesimal) canonical transformation that leaves given H invariant [29]. This implies that $\frac{dG}{dt} = \{G, H\} = \frac{\partial G}{\partial \mathbf{q}} \frac{\partial H}{\partial \mathbf{p}} - \frac{\partial G}{\partial \mathbf{p}} \frac{\partial H}{\partial \mathbf{q}} = 0$, where $\mathbf{p}$ and $\mathbf{q}$ are generalized positions and momenta. If $\nabla G = \vec{\mathbf{0}}$, then $\frac{dG}{dt}$ becomes zero regardless of the form of the Hamiltonian, and such a G represents a stationary transformation which conveys no information about the system. In this sense, spreading loss thus prompts the model to learn nontrivial canonical transformation by letting the model output from the set of noncanonical transformations, namely, the perturbed trajectory cannot form a constant of motion by a margin of Q.
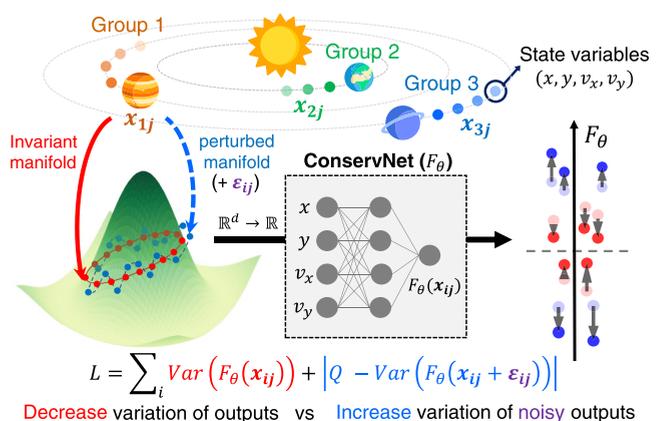
FIG. 1. Schematic overview of ConservNet and the role of noise-variance loss. Each group of data, which is a time series of planet trajectories in this example, is fed into the model and optimized to minimize the noise-variance loss.

III. NEURAL MODEL CONSTRUCTION AND TRAINING

ConservNet is a feedforward neural network constructed with four hidden layers with a layer width of 320 neurons and a single output neuron, using Mish [30] as an activation function. Our model receives system data $\mathbf{x}_{ij}$ and produces a single scalar value $F_\theta(\mathbf{x}_{ij})$ that aims to approximate the mapping function from states to conserved quantities. The noise vector $\boldsymbol{\varepsilon}_{ij}$ is newly sampled from the multivariate uniform distribution at every batch with the proper scaling. In practice, we employ standard deviation $\sigma(\mathbf{x}) = \sqrt{\text{Var}(\mathbf{x})}$ instead of variance $\text{Var}(\mathbf{x})$ as a measure of variance.

As a baseline for comparison, we trained a recently proposed Siamese neural network (SNN) [23] along with our model. This SNN architecture extracts an invariant by classifying whether two data points are from the same instance or not, similar to [21]. Both ConservNet and the SNN are trained with the Adam [31] optimizer using PYTORCH [32] for 50 000 epochs with early stoppings. For all experiments, $Q = 1$ and the spreading noise vector $\boldsymbol{\varepsilon}_{ij}$ is sampled from the uniform random vector with the maximum norm $R = 1$ [28].

IV. MODEL SYSTEMS AND DATA SETS

In this study, the ability of ConservNet is tested with three synthetic systems, two simulated model systems, and a real double-pendulum data set from [15]. The functional form of each invariant is presented in Table I. Three synthetic systems $S1$, $S2$, and $S3$ are constructed to show a variety of functional forms such as cubic, trigonometric, logarithmic, and rational functions. For the Lotka-Volterra system ($\frac{dx}{dt} = \alpha x - \gamma xy$ and $\frac{dy}{dt} = -\beta y + \delta xy$) [33] and the Kepler problem ($H_{\text{Kepler}} = \frac{\mathbf{p}^2}{2m} - \frac{GMm}{r}$), data are simulated by numerical integration with Euler's method. We find that normalizing the scale between variables improves performances, and thus variables with maximum values exceeding 10 are rescaled by a factor of 0.1 [28].

TABLE I. Systems and invariants for verification. We use $\alpha, \beta, \delta, \gamma = (1.1, 0.4, 0.1, 0.4)$ for the Lotka-Volterra system and $m = 1$ and $GM = 1$ for the Kepler problem. For the double-pendulum case, the ideal Hamiltonian is given.

System	Invariant formula
S1	$C = x_1 - 2x_2x_3 + 3x_4^2$
S2	$C = 3x_1 + 2\sin(x_2) + \sqrt{|x_1|}x_3^3$
S3	$C = 2x_1x_2 - [\ln(|x_1 + x_3|) - x_4]/x_3$
Lotka-Volterra	$C = \alpha \ln(x) + \delta \ln(y) - \beta x - \gamma y$
Kepler problem	$C_1 = xv_y - yv_x$
Kepler problem	$C_2 = \frac{1}{2}m(v_x^2 + v_y^2) - \frac{GMm}{r}$
Kepler problem	$C_3 = \mathbf{p} \times \mathbf{L} - mk\hat{\mathbf{r}}$
double-pendulum	$C_{\text{ideal}} = L_1^2(m_1 + m_2)\omega^2 + m_2L_2^2\omega^2$
(experiment)	$\quad + 2m_1m_2L_1L_2\omega_1\omega_2\cos(\theta_1 - \theta_2)$
	$\quad - 2gL_1(m_1 + m_2)\cos(\theta_1) - 2gm_2L_2\cos(\theta_2)$

V. RESULTS

We prepare 2000 training data with various data conditions (N, M) and an equal number of test data for all simulated systems, which is notable as a small amount compared to modern deep learning and other related studies [22–24] that typically employ more than 10 000 data. These conditions are addressed to replicate practical situations with high data costs and a limited number of different observations, common in physical and biological data. The code for data-set generation and model training is publicly available from [34].

The model performance of ConservNet is evaluated by the two aforementioned criteria: high correlation with the ground-truth invariant and small intragroup variance. We use Pearson correlation ρ and mean intragroup standard deviation $\bar{\sigma} = \frac{1}{N}\sum \sigma_i$ for each criterion.

Figure 2(a) illustrates the notable performances of ConservNet, simultaneously finds invariants from multiple groups at once, achieving strong Pearson correlation and small intragroup variation in every model system. For the case of multiple invariants in the Kepler problem, ConservNet captures the angular momentum first and finds the energy second when the angular momentum is controlled (see [28] for an analysis of multiple invariants). Figure 2(b) shows the resulting statistics of ConservNet for $S2(20, 100)$ as an example case. We can observe that our model shows smooth convergence without overfitting, while its $\bar{\sigma}$ decreases and ρ approaches 1. ConservNet shows consistent performance for different data conditions (N, M) as presented in [28].

We further investigate the capability and robustness of ConservNet by applying several different conditions prevalent in experimental data. First, we check the impact of noise on the data sets by adding noise $\mathcal{N}(0, s)$ with various strengths s. Figure 3(a) shows that ConservNet gives consistent performances under the noisy condition, with better correlation compared to the baseline. ConservNet effectively increases its data size by adding new random noise to its data set for each batch, thus showing better sample efficiency [35] and performance with inherent robustness to noise [36]. We also find that if the data have no invariant, our model draws attention to it by a strong overfitting [28].
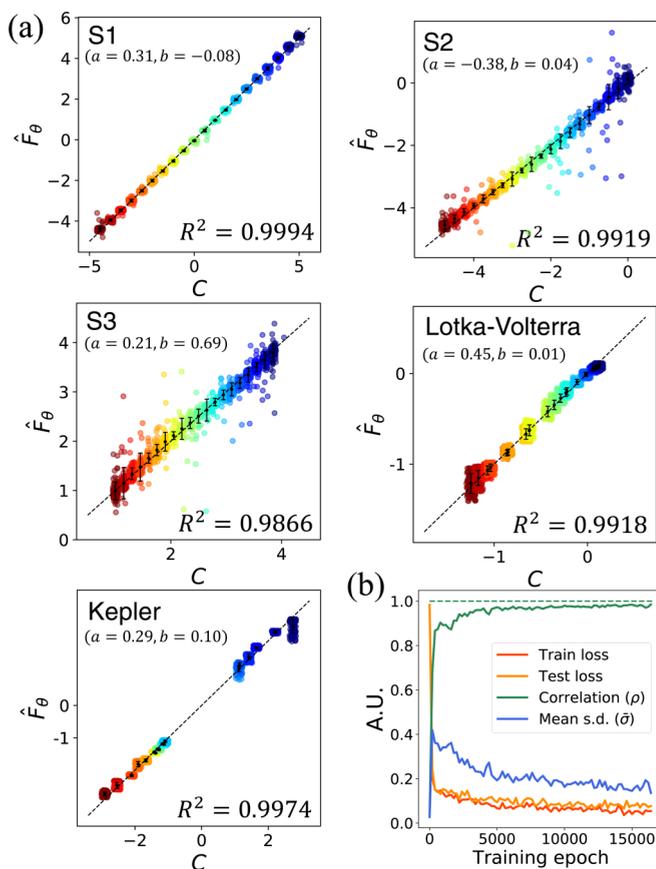
FIG. 2. Model performances of ConservNet. (a) Ground-truth invariants C versus fitted ConservNet outputs $\hat{F}_\theta = aF_\theta + b$ for $S1$, $S2$, $S3$, the Lotka-Volterra equation, and the Kepler problem are plotted under data condition $(20, 100)$ with R^2 values. Points with the same color share the same invariant values but are plotted at jittered values for visualization. The mean output value of each group (black dot) is shown with error bars for standard deviation and an identity line (dotted) is drawn for comparison. (b) Result statistics for invariant $S2(20, 100)$ with ideal correlation 1 (green dashed line).

In a real scenario, there might be irrelevant variables in an observed data set that do not compose the invariant. Filtering out such nuisance variables is crucial for data-driven
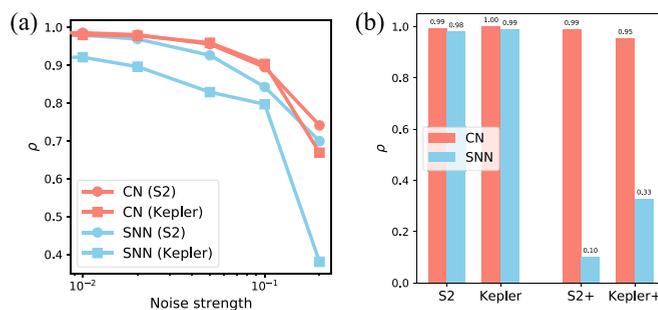
FIG. 3. Robustness of ConservNet (CN). (a) Pearson correlation of ConservNet and SNN for invariant $S2$ with various noise strengths. (b) Pearson correlation of ConservNet and SNN for two original data sets ($S2$ and Kepler) and their reinforced versions ($S2+$ and Kepler+) that include nuisance variables not appearing in the invariants.
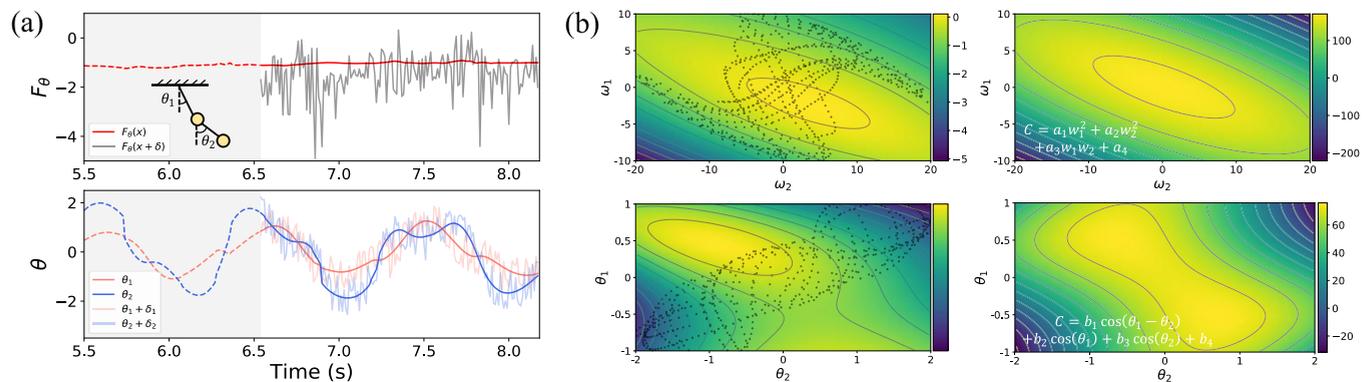
FIG. 4. ConservNet results for real double-pendulum data. (a) Model output $F_\theta(x)$ and the noisy model output $F_\theta(x + \boldsymbol{\varepsilon})$ with $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2)$ (top) and double-pendulum trajectories θ_1, θ_2 and noisy trajectories (bottom) versus time. Data in the shaded area are used for training (from 0 s to 6.54 s), with the remaining data used for testing (6.54 s to 8.18 s). (b) 2D heatmap of model output F_θ (left) and ideal Hamiltonian (right) for $(\theta_1, \theta_2, \omega_1, \omega_2) = (0, 0, \omega_1, \omega_2)$ (top) and $(\theta_1, \theta_2, \omega_1, \omega_2) = (\theta_1, \theta_2, 5, 10)$ (bottom). The training data points are scattered in the left panels, while the ideal formulas for the cross section are presented in the right panels. Here the ideal heatmaps are drawn with constants $(a_1, a_2, a_3, a_4 = 1, 0.32, 0.82, -170.95)$ and $(b_1, b_2, b_3, b_4 = 41, -124.13, -46, 82, 57)$, provided by [15].

discovery without any prior knowledge. We test our model with two reinforced data sets. First, we concatenate one extra variable $x_4 \sim \mathcal{N}(0, 1)$ to the $S2$ data set to construct $S2+$ with a noisy variable. Second, we transform the data of the Kepler problem from Cartesian coordinates $(x, \dot{x}, y, \dot{y})$ into polar coordinates $(r, \dot{r}, \theta, \dot{\theta})$ to construct Kepler+. In the polar coordinates, θ becomes a cyclic coordinate and neither $\dot{r}$ nor $\dot{\theta}$ appears in angular momentum $r\dot{\theta}$, different from the original Cartesian form $x\dot{y} - y\dot{x}$ where all of the state variables appear. As Fig. 3(b) shows, ConservNet exhibits robust performances even with the existence of the nuisance variables and coordinate transformation, while the SNN strongly overfits and shows low performance when there are unused variables, possibly due to the nature of classifiers and the absence of a proper regularizer.

Finally, we apply our model to a real double-pendulum trajectory from [15], which is a challenging task in a number of ways. According to [15], the data do not *strictly* obey any conservation laws due to noise and friction. Furthermore, the model has to discover the invariant in an extreme data condition where only a single trajectory ($N = 1$) with a limited number of data points ($M = 654$) is available for training. Note that the SNN is *inapplicable* to this case since it needs at least two groups of data to compare ($N \geqslant 2$).

We train our model and examine its output for stability and accuracy. Figure 4(a) shows that ConservNet output F_θ remains constant for the training and test trajectory but not for the noisy trajectory, verifying that ConservNet falls into neither trivial convergence nor overfitting and properly captures the functional form of the invariant. We further check two-dimensional cross sections of the model output by fixing two variables and varying two variables and compare them with the cross sections of the ideal four-dimensional Hamiltonian, constructed with the constants from [15]. The results are shown in Fig. 4(b). Considering inherent frictions and the restricted regions of the data points, both heatmaps are similar enough to the point where the inference of the abstract functional form is possible. To summarize, ConservNet successfully captured the conserved quantity from a real double-pendulum system with extreme data conditions.

VI. CONCLUSION AND OUTLOOK

In a real practice where the ground-truth invariant is unknown, we may identify the symbolic form of the invariant by sorting the output values and employing off-the-shelf polynomial regression or symbolic regression algorithms. We illustrate the result of such an application for invariant $S1$ as an example in [28], in which the ground-truth symbolic formula is successfully retrieved.

One limitation that ConservNet shares with [23] is that the single model finds a single invariant even if the system could have multiple invariants. While we showed that training with modified data leads to the discovery of remaining invariants, such modification is usually difficult for experimental data. Since our model identifies the numerical value of the invariant and [24] approximates the number of invariants, unifying the advantages of these approaches would be an interesting future direction to explore.

Also, ConservNet and all other attempts to solve the regression problem with neural networks are not free from extrapolation problems inherent to neural networks [37], as shown in the incomplete reconstruction of heatmap in Fig. 4(b). However, we note that the usual concern about this issue, generalization for unseen test data, does not completely hold to our model's main purpose. ConservNet and similar architecture for scientific discovery are mainly aimed at discovering scientific concepts from a given set of data rather than training on a specific datum and achieving good performance on a test datum (although we present the result from test data throughout this Letter, which demonstrates the model's intrapolation performance). Hence, if unseen data are newly collected, one may add such a data group into a cumulative data set and retrain ConservNet to fit the model for a potentially broader range of parameters.

In this Letter the invariants in a classical sense, such as a well-defined Hamiltonian, are mainly discussed. We can further expand the scope of ConservNet, for instance, by converting any nonautonomous system of x with multiple exogenous variables $y_1, y_2, \ldots, y_n$ to a standard form of $f(\frac{dx}{dt}, x, y_1, y_2, \ldots, y_n) = 0$, a system with an invariant

of value zero [38]. From this perspective, one can identify interactions among variables by discovering invariants with the proposed model. This opens a wide variety of potential applications of the model in academic disciplines where the underlying dynamics are yet to be discovered, including advanced domains of quantum mechanics [39,40], high-energy physics [41], astronomical science [42], and particle physics [43], where the scale of the data set is exceedingly large such that finding any meaningful structure is humanly intractable.

Automation of science with deep learning is a recently emerging field of study with plenty of uncharted research areas. The present work builds an interpretable connection between the data and scientists by extracting significant information from entangled high-dimensional data as a form of numerical value and symbolic equation, which can be further explained by a physicist. We envision that at some point, a neural network such as ConservNet or an integrated framework of such networks would automatically discover truly unseen knowledge from large-scale data sets.

ACKNOWLEDGMENTS

[1] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, Nat. Phys. **13**, 431 (2017).

[2] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Machine Learning Phases of Strongly Correlated Fermions, Phys. Rev. X **7**, 031038 (2017).

[3] E. P. Van Nieuwenburg, Y.-H. Liu, and S. D. Huber, Learning phase transitions by confusion, Nat. Phys. **13**, 435 (2017).

[4] Y. Zhang and E.-A. Kim, Quantum Loop Topography for Machine Learning, Phys. Rev. Lett. **118**, 216401 (2017).

[5] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[6] P. Baldi, P. Sadowski, and D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, Nat. Commun. **5**, 4308 (2014).

[7] P. Ponte and R. G. Melko, Kernel methods for interpretable machine learning of order parameters, Phys. Rev. B **96**, 205146 (2017).

[8] P. Zhang, H. Shen, and H. Zhai, Machine Learning Topological Invariants with Neural Networks, Phys. Rev. Lett. **120**, 066401 (2018).

[9] N. Sun, J. Yi, P. Zhang, H. Shen, and H. Zhai, Deep learning topological invariants of band insulators, Phys. Rev. B **98**, 085402 (2018).

[10] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).

[11] M. Rafayelyan, J. Dong, Y. Tan, F. Krzakala, and S. Gigan, Large-Scale Optical Reservoir Computing for Spatiotemporal Chaotic Systems Prediction, Phys. Rev. X **10**, 041037 (2020).

[12] J. L. Amey, J. Keeley, T. Choudhury, and I. Kuprov, Neural network interpretation using descrambler groups, Proc. Natl. Acad. Sci. **118**, e2016917118 (2021).

[13] R. D. King, J. Rowland, S. G. Oliver, M. Young, W. Aubrey, E. Byrne, M. Liakata, M. Markham, P. Pir, L. N. Soldatova *et al.*, The automation of science, Science **324**, 85 (2009).

[14] J. Bongard and H. Lipson, Automated reverse engineering of nonlinear dynamical systems, Proc. Natl. Acad. Sci. **104**, 9943 (2007).

[15] M. Schmidt and H. Lipson, Distilling free-form natural laws from experimental data, Science **324**, 81 (2009).

[16] E. Kaiser, J. N. Kutz, and S. L. Brunton, *Proceedings of the 57th IEEE Conference on Decision and Control, Miami Beach, 2018* (IEEE, Piscataway, 2018), pp. 6415–6421.

[17] T. Wu and M. Tegmark, Toward an artificial intelligence physicist for unsupervised learning, Phys. Rev. E **100**, 033311 (2019).

[18] H. Li, X.-q. Shi, M. Huang, X. Chen, M. Xiao, C. Liu, H. Chaté, and H. Zhang, Data-driven quantitative modeling of bacterial active nematics, Proc. Natl. Acad. Sci. **116**, 777 (2019).

[19] K. Champion, B. Lusch, J. N. Kutz, and S. L. Brunton, Data-driven discovery of coordinates and governing equations, Proc. Natl. Acad. Sci. **116**, 22445 (2019).

[20] R. Iten, T. Metger, H. Wilming, L. del Rio, and R. Renner, Discovering Physical Concepts with Neural Networks, Phys. Rev. Lett. **124**, 010508 (2020).

[21] A. Decelle, V. Martin-Mayor, and B. Seoane, Learning a local symmetry with neural networks, Phys. Rev. E **100**, 050102(R) (2019).

[22] Y.-i. Mototake, Interpretable conservation law estimation by deriving the symmetries of dynamics from trained deep neural networks, Phys. Rev. E **103**, 033303 (2021).

[23] S. J. Wetzel, R. G. Melko, J. Scott, M. Panju, and V. Ganesh, Discovering symmetry invariants and conserved quantities by interpreting siamese neural networks, Phys. Rev. Research **2**, 033499 (2020).

[24] Z. Liu and M. Tegmark, Machine Learning Conservation Laws from Trajectories, Phys. Rev. Lett. **126**, 180604 (2021).

[25] Y. Sun, *Deep Learning Face Representation by Joint Identification-Verification* (The Chinese University of Hong Kong, Hong Kong, 2015).

[26] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, in *37th International Conference on Machine Learning (ICML 2020), Vienna, Austria* (PMLR, 2020), pp. 1597–1607.

[27] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, Supervised contrastive learning, in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020*, virtual (2020).

[28] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.3.L042035 for the proof for the NV loss, model results with different constants and data conditions, application of symbolic regression, and details about dataset constructions.

[29] H. Goldstein, C. Poole, and J. Safko, *Classical Mechanics* (Addison Wesley, San Francisco, 2002).

[30] D. Misra, Mish: A self regularized non-monotonic neural activation function, arXiv:1908.08681.

[31] D. P. Kingma and J. Ba, in *3rd International Conference on Learning Representations, San Diego, 2015*, edited by Y. Bengio and Y. LeCun (ICLR, La Jolla, 2015).

[32] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, in *Advances in Neural Information Processing Systems 32*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran, Red Hook, 2019), pp. 8026–8037.

[33] Y. Takeuchi, *Global Dynamical Properties of Lotka-Volterra Systems* (World Scientific, Singapore, 1996).

[34] S. Ha and H. Jeong, nokpil/conservnet, http://doi.org/10.5281/zenodo.4491096 (CERN, Meyrin, 2021), version v1.0.0.

[35] Q. Xie, M.-T. Luong, E. Hovy, and Q. V. Le, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, Piscataway, 2020), pp. 10687–10698.

[36] R. G. Lopes, D. Yin, B. Poole, J. Gilmer, and E. D. Cubuk, Improving robustness without sacrificing accuracy with patch gaussian augmentation, arXiv:1906.02611 (2021).

[37] K. Xu, M. Zhang, J. Li, S. S. Du, K. Kawarabayashi and S. Jegelka, How neural networks extrapolate From feedforward to graph neural networks, in *9th International Conference on Learning Representations, Vienna, Austria*, edited by Y. Bengio and Y. LeCun (ICLR, La Jolla, 2021).

[38] P. E. Kloeden and M. Rasmussen, *Nonautonomous Dynamical Systems* (American Mathematical Society, Providence, 2011), Vol. 176.

[39] F. T. Hioe and J. H. Eberly, N-Level Coherence Vector and Higher Conservation Laws in Quantum Optics and Quantum Mechanics, Phys. Rev. Lett. **47**, 838 (1981).

[40] Y. Aharonov, S. Popescu, and D. Rohrlich, On conservation laws in quantum mechanics, Proc. Natl. Acad. Sci. **118**, e1921529118 (2021).

[41] N. H. Christ, Conservation-law violation at high energy by anomalies, Phys. Rev. D **21**, 1591 (1980).

[42] B. P. Abbott, R. Abbott, T. Abbott, M. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. Adhikari *et al.*, Observation of Gravitational Waves from a Binary Black Hole Merger, Phys. Rev. Lett. **116**, 061102 (2016).

[43] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. A. Khalek, A. A. Abdelalim, R. Aben, B. Abi, M. Abolins, O. AbouZeid *et al.*, Observation of a new particle in the search for the standard model Higgs boson with the ATLAS detector at the LHC, Phys. Lett. B **716**, 1 (2012).