

Mapping distinct phase transitions to a neural networkDimitrios Bachtis^{1,*}, Gert Aarts^{2,†} and Biagio Lucini^{1,3,‡}¹*Department of Mathematics, Swansea University, Bay Campus, SA1 8EN, Swansea, Wales, United Kingdom*²*Department of Physics, Swansea University, Singleton Campus, SA2 8PP, Swansea, Wales, United Kingdom*³*Swansea Academy of Advanced Computing, Swansea University, Bay Campus, SA1 8EN, Swansea, Wales, United Kingdom*

(Received 7 July 2020; accepted 22 October 2020; published 16 November 2020)

We demonstrate, by means of a convolutional neural network, that the features learned in the two-dimensional Ising model are sufficiently universal to predict the structure of symmetry-breaking phase transitions in considered systems irrespective of the universality class, order, and the presence of discrete or continuous degrees of freedom. No prior knowledge about the existence of a phase transition is required in the target system and its entire parameter space can be scanned with multiple histogram reweighting to discover one. We establish our approach in q -state Potts models and perform a calculation for the critical coupling and the critical exponents of the ϕ^4 scalar field theory using quantities derived from the neural network implementation. We view the machine learning algorithm as a mapping that associates each configuration across different systems to its corresponding phase and elaborate on implications for the discovery of unknown phase transitions.

DOI: [10.1103/PhysRevE.102.053306](https://doi.org/10.1103/PhysRevE.102.053306)**I. INTRODUCTION**

Deep learning [1] is a category of machine learning algorithms that has recently gained significant importance in the physical sciences. Applications of neural networks have emerged in research fields such as particle physics and cosmology, condensed matter physics, quantum computation, and physical chemistry. For a recent review see Ref. [2].

Within these developments machine learning has critically influenced the domain of statistical mechanics, particularly in the study of phase transitions [3,4]. A wide range of machine learning techniques, including neural networks [5–17], diffusion maps [18], support vector machines [19–22], and principal component analysis [23–27], have been implemented to study equilibrium and nonequilibrium systems. Transferable features have also been explored in phase transitions, including modified models through a change of lattice topology [3] or form of interaction [28], in Potts models with a varying odd number of states [29], in the Hubbard model [6], in fermions [5], in the neural network-quantum states ansatz [30,31], and in adversarial domain adaptation [32]. Among these, recent studies based on transferable features have mostly focused on predicting the critical temperature exclusively.

A machine learning algorithm can be implemented, based on a set of labeled training data, to complete a classification task such as the separation of phases in a statistical system. This is almost universally conducted under the assumption that the training and prediction data belong in identical probability distributions and have been acquired from the same

feature space. A change in feature space could arise by attempting to separate phases in a different system. However, such an attempt would require the algorithm to be rebuilt, due to the difference in systems, in order to solve, in essence, a new classification task. Transfer learning is a framework within the research field of machine learning introduced to address precisely this problem [33]. Specifically, it enables the transfer of knowledge from a solved machine learning task to a new one that shares certain similarities with the original.

In this paper, we propose transfer learning as a means to discover and study unknown phase transitions. In particular, we explore if the features learned by a convolutional neural network on the two-dimensional Ising model are sufficiently universal to predict the structure of phase transitions in systems irrespective of the universality class, the order, and the presence of discrete or continuous degrees of freedom. We further explore the deeper layers of the neural network architecture for universal features and discuss how the neural network acts as a mapping, which associates each configuration, across different systems under consideration, to its corresponding phase. It is then feasible to predict phases across a wide range of systems, governed by distinct Hamiltonians and symmetries, without introducing prior knowledge about the presence of a phase transition in them.

To study unknown phase transitions, we treat the predictive function of the Ising-trained neural network as an observable in the target system, and propose the extension of previous work on reweighting for machine learning [34], to the multiple histogram method [35,36]. This technique enables the accurate definition of the target system's critical region by interpolating the neural network function in its entire parameter space. Given the knowledge of an effective order parameter in a target system through transfer learning, we train a randomly initialized neural network to perform a finite-size scaling analysis based on histogram-reweighted quantities derived

*dimitrios.bachtis@swansea.ac.uk

†g.aarts@swansea.ac.uk

‡b.lucini@swansea.ac.uk

from the machine learning algorithm. This results in an accurate calculation of multiple critical exponents and the critical coupling.

To establish our approach, we apply the Ising-trained neural network to q -state Potts models, systems that manifest first- or second-order phase transitions depending on their number of states, and the ϕ^4 scalar field theory. Without using knowledge about the presence of a phase transition in the target system, we employ transfer learning to reconstruct an effective order parameter therein, define rigidly the boundaries of its critical region, and perform a finite-size scaling analysis on machine learning derived quantities to determine its universality class.

II. TRANSFER LEARNING

We consider a domain Δ which comprises a feature space Ω and a marginal probability distribution $P(X)$, where $X = \{x_i, \dots, x_N\} \in \Omega$ is a learning sample:

$$\Delta = \{\Omega, P(X)\}. \quad (1)$$

For the case of the Ising model, the feature space Ω contains all possible configurations of the system, X is a subset of Ω , comprising a finite number of configurations that have been drawn from selected inverse temperatures, and x_i corresponds to a specific configuration.

Within a domain Δ , we define a task T , with a label space Y and a predictive function $f(\cdot)$ which is learned during the optimization of the machine learning algorithm on the training data:

$$T = \{Y, f(\cdot)\}. \quad (2)$$

In particular, for a phase identification task in the Ising model, a label $y_i \in Y$, denoted by zero or one, signifies the disordered or the ordered phase, and the predictive function $f(\cdot)$, which is equivalent to the conditional probability distribution $p(y|x)$, associates each configuration x , given as input to the algorithm, to its corresponding phase.

Given the knowledge acquired on a source domain Δ_s and learning task T_s , transfer learning can be utilized for a target domain Δ_t and learning task T_t , to enhance the predictive function $f(\cdot)_t$, when $\Delta_s \neq \Delta_t$ or $T_s \neq T_t$ [33]. The condition $\Delta_s \neq \Delta_t$ implies, based on Eq. (1), that the feature spaces or the marginal probability distributions might be different: $\Omega_s \neq \Omega_t$, $P_s(X) \neq P_t(X)$. An example of domain adaptation concerns employing transfer learning to relocate from a source domain of a two-dimensional binary system, such as the Ising model, to a different system in search for a phase transition that separates a disordered from an ordered phase.

III. DISCOVERING PHASE TRANSITIONS

We employ transfer learning to predict the phase diagram of two-dimensional q -state Potts models and the ϕ^4 scalar field theory using an Ising-trained convolutional neural network and multiple histogram reweighting. In addition we explore if the neural network accurately classifies distinct phase transitions due to the presence of learned universal features in deeper layers.

The configurations of the Ising and the Potts models are obtained using Markov chain Monte Carlo simulations with the Wolff algorithm [37]. The scalar field theory is simulated with the Metropolis algorithm followed by a sweep of the Wolff algorithm [38,39] (see Appendix A for details of the models). The convolutional neural network (see Fig. 1) is trained on the Ising model, where configurations have binary degrees of freedom, mapped to -1 and 1 . Training of the neural network is conducted for $\beta \leq 0.41$ and $\beta \geq 0.47$ in the disordered and ordered phases, respectively (see Appendix B for details of the neural network). To be consistent with the same range when conducting transfer learning, configurations from Potts models are mapped to unique positive and negative numbers between -1 and 1 . By choosing unique values for the states of the Potts model the physics of the Hamiltonian, which includes a δ function, is retained. When the number of states is odd, the remaining value is chosen arbitrarily as positive or negative. The degrees of freedom for the ϕ^4 scalar field theory lie within $-\infty < \phi < \infty$ and no normalization is conducted.

Once the convolutional neural network is trained on the source domain Δ_s , to classify the ordered and disordered phases of the Ising model, its predictive function $f_s(\cdot)$, which is interpreted here as the probability of being in the ordered phase, is employed to predict the phase of a configuration x . The application of the predictive function to an importance-sampled configuration x converts $f_s(x)$ into an observable with an attached Boltzmann weight, enabling its extrapolation with reweighting in the statistical system's parameter space [34]. Here we propose the extension of reweighting to the multiple histogram method (see Appendix C for a derivation). This technique combines an arbitrary number of simulated data sets and enables the estimation of the partition function, and therefore of the predictive function, in the system's entire parameter space [35].

The results can be seen in Fig. 2(a), where the predictive function, depicted by the line, is obtained in the visible range using multiple histogram reweighting on a combination of Monte Carlo data sets. We recall that the CNN is trained on the intervals $\beta \leq 0.41$ and $\beta \geq 0.47$. The results are compared with calculations of the neural network on independent Monte Carlo simulations, which lie within statistical errors. The predictive function acts as an effective order parameter, and the use of multiple histogram reweighting advances previous work by enabling the estimation of the predictive function over extended ranges in the system's parameter space.

The predictive function $f_s(\cdot)$ was learned on configurations of the Ising model, but the classification knowledge can be adapted to other domains by applying it to configurations x' of a different system. The predictive function $f_s(x')$ then becomes an observable in the target system, where multiple histogram reweighting can be employed to estimate it in the target parameter space. The results for Potts models can be seen in Fig. 2(c) where we recall that the phase transition is second-order for $q = 3, 4$ and first-order for $q = 5, 6, 7$ [40]. The dashed vertical lines indicate the positions of the analytically determined critical coupling, $\beta_c = \ln(1 + \sqrt{q})$. Good agreement between the results obtained by transfer learning and the exact ones can already be observed on the volume shown here ($L = 128$). The results for the ϕ^4 scalar field

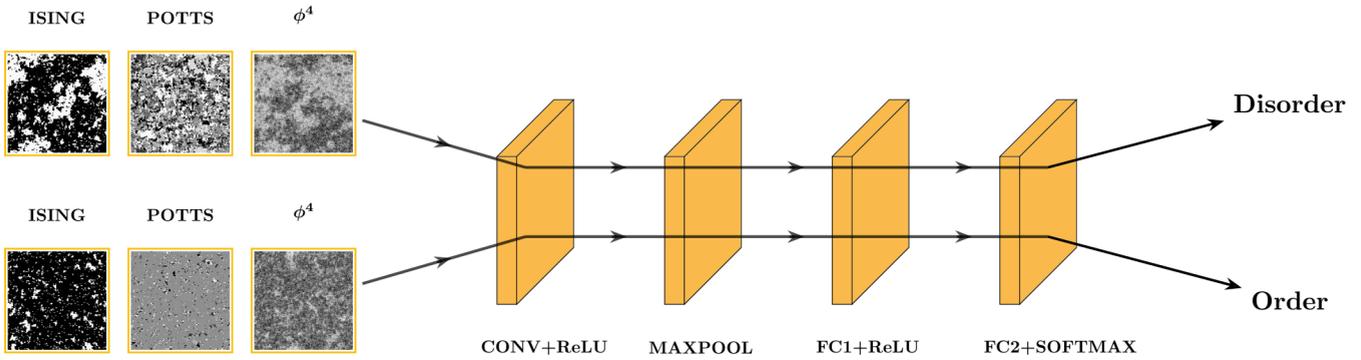


FIG. 1. The 2D Ising-trained convolutional neural network (see Appendix B). Configurations are successfully classified as belonging to the disordered or ordered phase, irrespective of the system.

theory, a system with continuous degrees of freedom, can be seen in Fig. 2(b), where we have fixed the dimensionless quartic coupling $\lambda_L = 0.7$ and reweighted based on the values of μ^2 , the dimensionless mass parameter (see Appendix A). We infer that the system is in the broken (ordered) phase for large and negative μ^2 and in the symmetric (disordered) phase for smaller μ^2 values. In Sec. IV we analyze the critical properties of the scalar field theory further using a finite-size scaling analysis.

We note that the neural network successfully differentiates between ordered and disordered phases, irrespective of the

system, and despite changes in discrete or continuous degrees of freedom, the universality class, and the order of the phase transition. Consequently, the predictive function $f_s(\cdot)$ learned by a convolutional neural network on the two-dimensional Ising model is capable to reconstruct effective order parameters in more complicated systems.

To gain further insights about the capability of the neural network to predict phases across different systems, we consider that the fundamental constituents of a trained neural network are the sets of variational parameters, specifically the weights and the biases at each layer of its architecture.

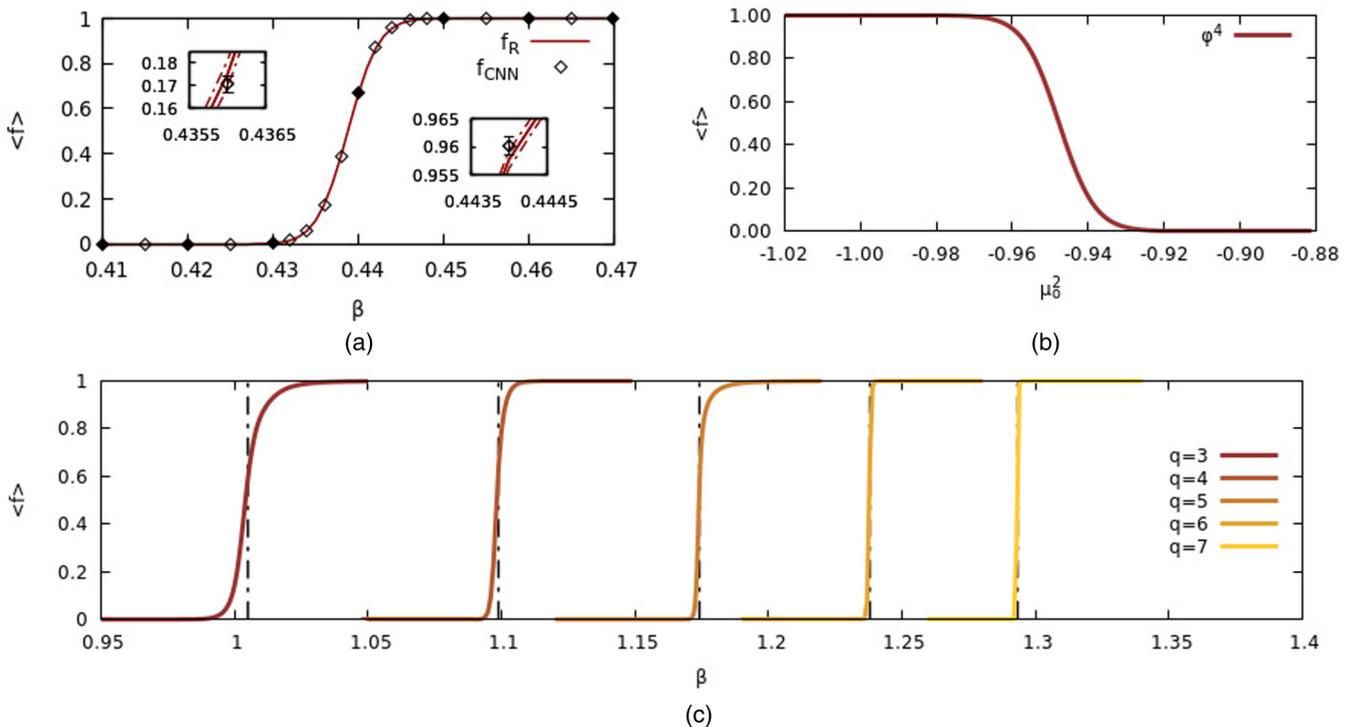


FIG. 2. Multiple histogram reweighting of the predictive function versus coupling of the two-dimensional (a) Ising model, (b) ϕ^4 scalar field theory and (c) $q = 3, \dots, 7$ (left to right) Potts models, for lattice size $L = 128$. The reweighted predictive function is depicted by the line. Its statistical errors are smaller than its width, excluding the insets in (a), where errors are portrayed by dashed lines. The results for all figures have been obtained based on a convolutional neural network trained exclusively on the Ising model. In (a) filled points correspond to the combined Monte Carlo data sets used to conduct reweighting and empty points are calculations of the predictive function on independent Monte Carlo data sets, added for comparison. In (c) vertical dashed lines have been added for the values of critical inverse temperature of Potts models, calculated through Eq. (A2).

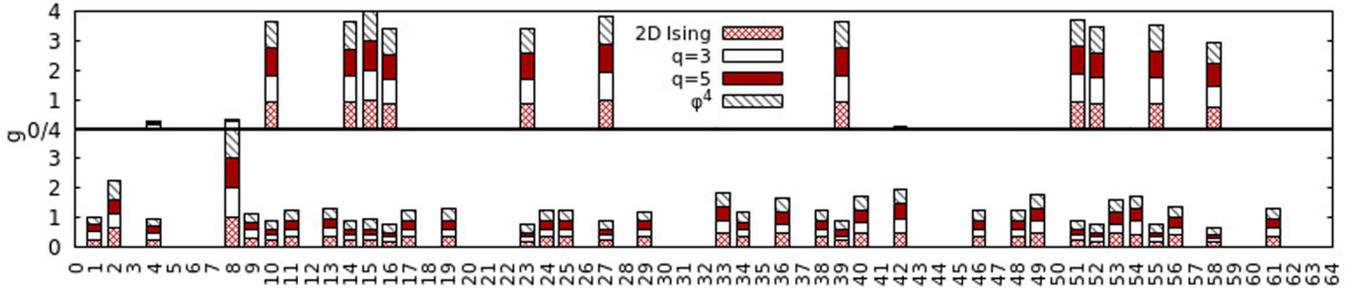


FIG. 3. Mean activations g versus the 64 latent variables in the fully connected layer (FC1) of the 2D Ising-trained convolutional neural network. Results are from the two-dimensional Ising model, $q = 3$ and $q = 5$ Potts model, and the ϕ^4 field theory for configurations belonging in the disordered (top) and ordered phase (bottom). The maximum values have been normalized to one, and the results have been rescaled accordingly and stacked vertically for better representation.

These are optimized during the training process for a specific machine learning task T . The variational parameters eventually converge to certain values which encode the solution to the particular problem under consideration. In our approach they have been tuned to accurately separate the disordered and ordered phases of the two-dimensional Ising model based on a set of labeled Ising configurations. It is well known that within the first layers of a neural network architecture the variational parameters correspond to learned universal features [41]. This form of universality is expected to eventually diminish towards deeper layers of the neural network architecture where the features are anticipated to transition from universal to specific in relation to the machine learning task T .

Considering that the 2D Ising-trained neural network successfully classifies phases across different systems by utilizing an identical predictive function $f_s(\cdot)$, we expect universal features to extend further in deeper layers. The activation function of a variable in an intermediate layer of the neural network architecture acts as a transformation that maps a certain input to an output representation. For a feature to be deemed universal in relation to certain inputs, the corresponding activation function should produce identical output representations. We therefore present as input to the neural network configurations from different systems and calculate their mean activations in order to diminish statistical errors during classification. The results are depicted in Fig. 3, where the activations have been drawn for the 64 variables of the first fully connected layer. We observe spikes for values of certain activations, which are consistent for configurations from the ordered or disordered phase, irrespective of the system. The features learned by the Ising-trained convolutional neural network successfully map each configuration, across systems under consideration, to its associated phase.

IV. STUDYING UNKNOWN PHASE TRANSITIONS

The preceding results designate that a neural network trained on a prototypical system, such as the Ising model which manifests a second-order phase transition, can serve as a tool to discover phase transitions in more complicated systems. This pursuit is greatly enhanced with the use of multiple histogram reweighting, where the entire parameter space can be examined for the discovery of a phase transition. By obtaining the knowledge of an effective order parameter

through the use of transfer learning in a target system, the boundaries of its critical region can then be accurately defined.

As an example, for the ϕ^4 field theory with fixed $\lambda_L = 0.7$, we presume based on Fig. 2(b), that a conservative choice for the boundaries of the ordered and disordered phases lie on values of bare mass $\mu^2 \leq -1.0$ and $\mu^2 \geq -0.90$, respectively. We note that one could employ the neural network that produced the results of Fig. 2(b) to study the infinite-volume limit of the target system. However, since target systems might differ from the source system in fundamental aspects such as the order of the phase transition, the continuous degrees of freedom or the underlying symmetries, we treat the preceding results from transfer learning as qualitative indications. To obtain quantitative results, we train a randomly initialized neural network on the ϕ^4 scalar field theory to study the infinite volume limit. We sample configurations which are labeled as belonging to each phase based on Fig. 2(b). For the randomly initialized neural network, which is trained using the same setup (see Appendix B), the source domain is the one defined by the configurations of ϕ^4 scalar field theory, and the machine learning task T is the separation of the two phases, which are discovered through transfer learning.

We proceed by performing a calculation of the critical μ_c^2 and the critical exponents of the ϕ^4 scalar field theory under a finite size scaling assumption, by relying on quantities derived from the neural network implementation and their reweighted extrapolations [34]. Specifically, we associate pseudocritical points based on the maxima of the fluctuations of the predictive function:

$$\delta f(\cdot) = V(\langle f^2 \rangle - \langle f \rangle^2), \tag{3}$$

where V is the volume of the system. The maxima (see Table I) are obtained by reweighting based on one Monte Carlo data set within the critical region. We note that including measurements from additional simulations in reweighting always leads to a reduction in statistical errors [35]. We then calculate the correlation length critical exponent and critical μ_c^2 through the equation

$$\left| \frac{\mu_c^2(L) - \mu_c^2}{\mu_c^2} \right| \sim \xi^{-\frac{1}{\nu}} \sim L^{-\frac{1}{\nu}}, \tag{4}$$

TABLE I. Pseudocritical points $\mu_c^2(L)$ for fixed $\lambda_L = 0.7$ and maxima of the predictive function δf_{\max} for various lattice sizes L of the ϕ^4 scalar field theory.

L	$\mu_c^2(L)$	δf_{\max}
200	-0.94988(4)	8239(50)
256	-0.95037(5)	12 915(56)
360	-0.95096(4)	22 348(138)
440	-0.95117(3)	34 710(211)

and the magnetic susceptibility exponent through the equation

$$\delta f \sim L^{\frac{\gamma}{\nu}}. \tag{5}$$

The results are presented in Figs. 4 and 5 and in Table II. We note that the value of the critical μ_c^2 for fixed $\lambda_L = 0.7$ is within statistical errors from calculations of conventional phase transition indicators, such as the susceptibility or the Binder cumulant, in Refs. [38,42]. For the values of λ_L and μ^2 , we anticipate the phase transition to be in the universality class of the two-dimensional Ising model, as evidenced by the calculation of the critical exponents (see Appendix A). In the error analysis (see Appendix D), only statistical errors from predictions on a finite Monte Carlo data set were considered. Better accuracy can be achieved by including calculations from additional lattice sizes.

V. CONCLUSIONS

In this paper we employed transfer learning to discover and study phase transitions from ordered to disordered phases in two-dimensional statistical systems. By employing an Ising-trained convolutional neural network and multihistogram reweighting, we reconstructed effective order parameters in q -state Potts models and the ϕ^4 scalar field theory, without introducing prior knowledge about the presence of a phase transition in those target systems. In addition we viewed the neural network as a mapping that associates each configuration, across different systems, to its corresponding phase and uncovered universal features learned on deeper layers of the neural network architecture. Furthermore we utilized transfer learning to define the boundaries of the critical region in the ϕ^4 scalar field theory. We finally conducted a finite size scaling analysis to calculate multiple critical exponents and the critical μ_c^2 constant using quantities derived solely from the neural network algorithm. We are not aware of any prior work

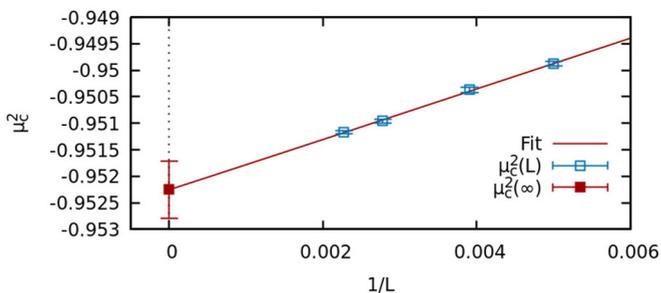


FIG. 4. Pseudocritical μ_c^2 versus inverse lattice size.

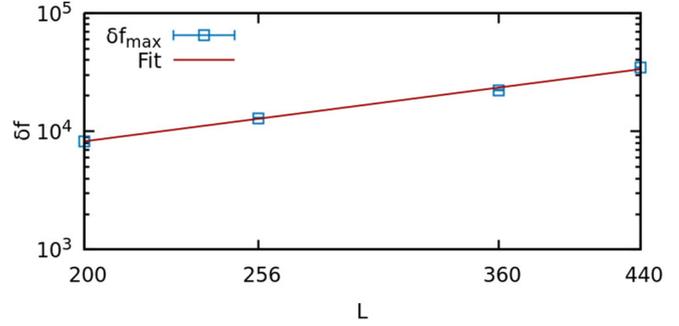


FIG. 5. Fluctuations of the predictive function versus lattice size on double logarithmic scale.

that employs machine learning to calculate critical exponents for a quantum field theory.

The discovery and study of an unknown phase transition can be completed in four steps:

Firstly, a neural network is trained on labeled configurations of an original system to learn a predictive function that separates a number of phases.

Secondly, the predictive function is then applied on unlabeled configurations of a target system, to recognize if they belong to any of the learned phases. The entire parameter space can be inspected with multiple histogram reweighting to uncover the target system’s critical region.

Thirdly, given this knowledge a second neural network is trained on configurations of the target system that have been labeled in compliance with the recognized phases in the previous step.

Finally, the infinite volume limit is studied, using reweighting on derived quantities from the second neural network implementation to determine the universality class of the phase transition in the target system.

Transfer learning enables the use of simplistic systems to study complicated models with partially known behavior. Combined with multihistogram reweighting, a technique that is able to scan the entire parameter space by reconstructing effective order parameters therein, it permits the discovery of unknown phase transitions. Transfer learning is employed successfully to predict the phase structure of target systems irrespective of the universality class, order, and discrete or continuous degrees of freedom, and can therefore be employed to uncover phase transitions in seemingly unrelated systems. Finally one could combine multiple transfer learning implementations, obtained by training neural networks on models with different phase transitions, to predict the structure of an intricate, unknown phase transition in a target system.

Note added: Recently Refs. [43,44] appeared on the arXiv. Reference [43] employs transfer learning for Potts models,

TABLE II. Critical μ_c^2 for fixed $\lambda_L = 0.7$ and critical exponents of the ϕ^4 scalar field theory.

	μ_c^2	ν	γ/ν
CNN+Reweighting	-0.95225(54)	0.99(34)	1.78(7)

and Ref. [44] explores universal critical behavior of phase transitions based on the intrinsic dimension in the data space.

ACKNOWLEDGMENTS

The authors received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement No. 813942. The work of G.A. and B.L. has been supported in part by the STFC Consolidated Grant No. ST/P00055X/1. The work of B.L. is further supported in part by the Royal Society Wolfson Research Merit Award No. WM170010 and by the Leverhulme Foundation Research Fellowship RF-2020-46119. Numerical simulations have been performed on the Swansea SUNBIRD system. This system is part of the Supercomputing Wales project, which is partly funded by the European Regional Development Fund (ERDF) via the Welsh Government. We thank COST Action CA15213 THOR for support.

APPENDIX A: ISING, POTTS AND THE ϕ^4 FIELD THEORY

We consider the q -state Potts model on a square lattice, with a Hamiltonian:

$$H_P = -J \sum_{\langle ij \rangle} \delta(\sigma_i, \sigma_j), \quad (\text{A1})$$

where J denotes the coupling constant, which we set to one, $\langle ij \rangle$ is a sum over nearest-neighbor interactions, $\delta(\sigma_i, \sigma_j)$ is the Kronecker delta and σ_i is a spin at lattice site i that can take the values $1, \dots, q$. The 2D Potts model has a second-order phase transition for $q \leq 4$ and a first-order phase transition for $q \geq 5$ [40]. For $q = 2$, $\sigma \in \{-1, 1\}$, $J_{\text{ising}} = J/2$, the Potts model reduces to the Ising model. The inverse critical temperature of the 2D q -state Potts model is given by

$$\beta_c^{\text{Potts}} = \ln(1 + \sqrt{q}), \quad (\text{A2})$$

where for $q = 2$, $\beta_c^{\text{ising}} = \beta_c^{\text{Potts}}/2$.

We consider the ϕ^4 scalar field theory in two dimensions, described by the Euclidean Lagrangian

$$\mathcal{L}_E = \frac{1}{2}(\nabla\phi)^2 + \frac{1}{2}\mu_0^2\phi^2 + \frac{\lambda}{4}\phi^4. \quad (\text{A3})$$

We discretize the ϕ^4 scalar field theory on a square lattice with lattice spacing a :

$$S_E = \sum_n \left[\frac{1}{2} \sum_{v=1}^{d=2} (\phi_{n+e_v} - \phi_n)^2 + \frac{1}{2} \mu^2 \phi_n^2 + \frac{1}{4} \lambda_L \phi_n^4 \right],$$

where the dimensionless parameters $\mu^2 = a^2 \mu_0^2$ and $\lambda_L = a^2 \lambda$ are the (bare) mass squared and coupling constant. By fixing the value of λ_L while varying μ^2 we cross a second-order phase transition in the (λ_L, μ^2) plane [38]. When $\lambda_L \rightarrow \infty$ and $\mu^2 \rightarrow -\infty$ the system is expected to be in the same universality class as the 2D Ising model, whereas for λ_L fixed and $\mu \rightarrow 0$ the emerging critical behavior is Gaussian [45].

APPENDIX B: CONVOLUTIONAL NEURAL NETWORK

The neural network architecture, implemented with TensorFlow and the Keras library, comprises a convolutional layer

with 64 filters, of size 2×2 and a stride of $s = 2$. The result is then forwarded to a max-pooling layer of size 2×2 and subsequently to a fully connected layer (FC1) with 64 latent variables. The nonlinear function is chosen as a rectified linear unit $g(x) = \max(0, x)$. The output layer consists of a fully connected layer (FC2) with a softmax activation function and the training is conducted with the Adam algorithm and a mini batch size of 12. The architecture was optimized for the Ising model based on the training and validation loss.

The neural network is trained on 1000 uncorrelated configurations per each parameter choice. Specifically for the case of the Ising model the range of inverse temperatures is $0.32, \dots, 0.41$ and $0.47, \dots, 0.56$ in the disordered and the ordered phases, respectively, with a step size of 0.01. The second neural network is trained independently on the ϕ^4 field theory on values of bare mass $-1.09, \dots, -1.00$ and $-0.90, \dots, -0.81$ with the same step size.

APPENDIX C: MULTIHISTOGRAM REWEIGHTING

During a Monte Carlo simulation the probability of generating a state with a lattice action (or energy) S is

$$p(S) = \rho(S) \frac{\exp[-\sum_k g^{(k)} S^{(k)}]}{Z}, \quad (\text{C1})$$

where $\rho(S)$ the density of states, Z the partition function, and S a lattice action (or Hamiltonian) that separates in terms of a set of parameters $\{g^{(k)}\}$:

$$S = \sum_k g^{(k)} S^{(k)}. \quad (\text{C2})$$

The probability $p(S)$ can be estimated during a Monte Carlo simulation through $p(S) = N(S)/n$ where n is the number of independent measurements and $N(S)$ the histograms of the action. After conducting a series of simulations for a specific set of parameters $\{g_i^{(k)}\}$, we arrive at a number of different estimates ρ_i of the density of states, given by

$$\rho_i(S) = \frac{N_i(S) Z_i}{n_i \exp[-\sum_k g_i^{(k)} S^{(k)}]}. \quad (\text{C3})$$

The estimates can be combined through a weighted average to estimate the density of states:

$$\rho(S) = \sum_i w_i \rho_i(S). \quad (\text{C4})$$

The values w_i are acquired through a minimization of the variance in the density of states, resulting in the equation

$$\rho(S) = \frac{\sum_i N_i(S)}{\sum_j n_j Z_j^{-1} \exp[-\sum_k g_j^{(k)} S^{(k)}]}, \quad (\text{C5})$$

where j is the number of available Monte Carlo data sets. For each data set, simulated on a parameter set $\{g_m^{(k)}\}$, the partition function, given by the equation

$$Z_m = \sum_S \rho(S) \exp\left[-\sum_k g_m^{(k)} S^{(k)}\right] \quad (\text{C6})$$

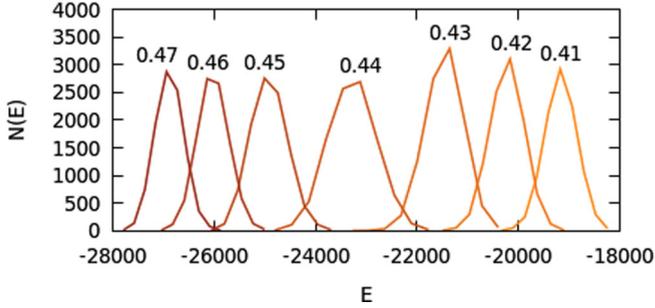


FIG. 6. Histograms of energies, from Monte Carlo data sets simulated at inverse temperatures $\beta = 0.41, \dots, 0.47$ of the 2D Ising model for $L = 128$. Overlapping histograms enable the interpolation over the entire inverse temperature range with multiple histogram reweighting.

can be estimated through an iterative scheme [35,46], by solving

$$Z_m = \sum_{i,s} \frac{1}{\sum_j n_j Z_j^{-1} \exp \left[\sum_k (g_m^{(k)} - g_j^{(k)}) S_{is}^{(k)} \right]}. \quad (\text{C7})$$

After convergence, the partition function Z_l for an interpolated parameter set $\{g_l^{(k)}\}$ can be acquired by one iteration of Eq. (C7). The expectation value of an arbitrary observable

$\langle O \rangle_l$ for a set of parameters $\{g_l^{(k)}\}$ is then calculated using the equation

$$\langle O \rangle_l = \frac{1}{Z_l} \sum_{i,s} \frac{O_{is}}{\sum_j n_j Z_j^{-1} \exp \left[\sum_k (g_l^{(k)} - g_j^{(k)}) S_{is}^{(k)} \right]}, \quad (\text{C8})$$

where s is the number of states sampled at i_{th} simulation. By carefully conducting simulations in order to obtain overlapping histograms between parameter values (e.g., see Fig. 6) one can calculate the partition function and arbitrary observables of interest in the entire parameter space using Eqs. (C7) and (C8).

APPENDIX D: BINNING ERROR ANALYSIS

The error analysis is conducted with a binning approach. A Monte Carlo data set, comprising 10 000 uncorrelated measurements, is separated into $n_b = 10$ groups of 1000 measurements. Multihistogram reweighting is conducted n_b times for extrapolated sets of parameters using the measurements of each group in each available simulation. The standard deviation for the expectation value at each extrapolated parameter is given through the equation

$$\sigma = \sqrt{\frac{1}{n_b - 1} (\overline{O^2} - \overline{O}^2)}. \quad (\text{D1})$$

-
- [1] I. J. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, USA, 2016), <http://www.deeplearningbook.org>.
- [2] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* **91**, 045002 (2019).
- [3] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, *Nat. Phys.* **13**, 431 (2017).
- [4] E. L. van Nieuwenburg, Y.-H. Liu, and S. Huber, Learning phase transitions by confusion, *Nat. Phys.* **13**, 435 (2017).
- [5] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, Machine learning quantum phases of matter beyond the fermion sign problem, *Sci. Rep.* **7**, 8823 (2017).
- [6] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Machine Learning Phases of Strongly Correlated Fermions, *Phys. Rev. X* **7**, 031038 (2017).
- [7] A. Tanaka and A. Tomiya, Detection of phase transition via convolutional neural networks, *J. Phys. Soc. Jpn.* **86**, 063001 (2017).
- [8] G. Torlai and R. G. Melko, Learning thermodynamics with Boltzmann machines, *Phys. Rev. B* **94**, 165134 (2016).
- [9] P. Suchsland and S. Wessel, Parameter diagnostics of phases and phase transition learning by neural networks, *Phys. Rev. B* **97**, 174435 (2018).
- [10] Q. Ni, M. Tang, Y. Liu, and Y.-C. Lai, Machine learning dynamical phase transitions in complex networks, *Phys. Rev. E* **100**, 052312 (2019).
- [11] Y.-T. Hsu, X. Li, D.-L. Deng, and S. Das Sarma, Machine Learning Many-Body Localization: Search for the Elusive Non-ergodic Metal, *Phys. Rev. Lett.* **121**, 245701 (2018).
- [12] E. van Nieuwenburg, E. Bairey, and G. Refael, Learning phase transitions from dynamics, *Phys. Rev. B* **98**, 060301(R) (2018).
- [13] J. Venderley, V. Khemani, and E.-A. Kim, Machine Learning Out-of-Equilibrium Phases of Matter, *Phys. Rev. Lett.* **120**, 257204 (2018).
- [14] C. Alexandrou, A. Athenodorou, C. Chrysostomou, and S. Paul, The critical temperature of the 2D-Ising model through Deep Learning Autoencoders, [arXiv:1903.03506](https://arxiv.org/abs/1903.03506) [cond-mat.stat-mech].
- [15] M. N. Chernodub, H. Erbin, V. A. Goy, and A. V. Molochkov, Topological defects and confinement with machine learning: The case of monopoles in compact electrodynamics, *Phys. Rev. D* **102**, 054501 (2020).
- [16] K. Zhou, G. Endrődi, L.-G. Pang, and H. Stöcker, Regressive and generative neural networks for scalar field theory, *Phys. Rev. D* **100**, 011501(R) (2019).
- [17] K. A. Nicoli, S. Nakajima, N. Strodthoff, W. Samek, K.-R. Müller, and P. Kessel, Asymptotically unbiased estimation of physical observables with neural samplers, *Phys. Rev. E* **101**, 023304 (2020).
- [18] J. F. Rodriguez-Nieva and M. S. Scheurer, Identifying topological order through unsupervised machine learning, *Nat. Phys.* **15**, 790 (2019).
- [19] P. Ponte and R. G. Melko, Kernel methods for interpretable machine learning of order parameters, *Phys. Rev. B* **96**, 205146 (2017).
- [20] C. Giannetti, B. Lucini, and D. Vadacchino, Machine learning as a universal tool for quantitative investigations of phase transitions, *Nucl. Phys. B* **944**, 114639 (2019).

- [21] J. Greitemann, K. Liu, and L. Pollet, Probing hidden spin order with interpretable machine learning, *Phys. Rev. B* **99**, 060404(R) (2019).
- [22] K. Liu, J. Greitemann, and L. Pollet, Learning multiple order parameters with interpretable machines, *Phys. Rev. B* **99**, 104410 (2019).
- [23] L. Wang, Discovering phase transitions with unsupervised learning, *Phys. Rev. B* **94**, 195105 (2016).
- [24] W. Hu, R. R. P. Singh, and R. T. Scalettar, Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination, *Phys. Rev. E* **95**, 062122 (2017).
- [25] C. Wang and H. Zhai, Machine learning of frustrated classical spin models. I. Principal component analysis, *Phys. Rev. B* **96**, 144432 (2017).
- [26] N. C. Costa, W. Hu, Z. J. Bai, R. T. Scalettar, and R. R. P. Singh, Principal component analysis for fermionic critical points, *Phys. Rev. B* **96**, 195138 (2017).
- [27] W. Zhang, L. Wang, and Z. Wang, Interpretable machine learning study of the many-body localization transition in disordered quantum Ising spin chains, *Phys. Rev. B* **99**, 054208 (2019).
- [28] A. Canabarro, F. F. Fanchini, A. L. Malvezzi, R. Pereira, and R. Chaves, Unveiling phase transitions with machine learning, *Phys. Rev. B* **100**, 045129 (2019).
- [29] K. Shiina, H. Mori, Y. Okabe, and H. K. Lee, Machine-learning studies on spin models, *Sci. Rep.* **10**, 2177 (2020).
- [30] R. Zen, L. My, R. Tan, F. Hébert, M. Gattobigio, C. Miniatura, D. Poletti, and S. Bressan, Transfer learning for scalability of neural-network quantum states, *Phys. Rev. E* **101**, 053301 (2020).
- [31] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [32] P. Huembeli, A. Dauphin, and P. Wittek, Identifying quantum phase transitions with adversarial neural networks, *Phys. Rev. B* **97**, 134109 (2018).
- [33] S. J. Pan and Q. Yang, A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.* **22**, 1345 (2010).
- [34] D. Bachtis, G. Aarts, and B. Lucini, Extending machine learning classification capabilities with histogram reweighting, *Phys. Rev. E* **102**, 033303 (2020).
- [35] A. M. Ferrenberg and R. H. Swendsen, Optimized Monte Carlo Data Analysis, *Phys. Rev. Lett.* **63**, 1195 (1989).
- [36] A. M. Ferrenberg and R. H. Swendsen, New Monte Carlo Technique for Studying Phase Transitions, *Phys. Rev. Lett.* **61**, 2635 (1988).
- [37] U. Wolff, Collective Monte Carlo Updating for Spin Systems, *Phys. Rev. Lett.* **62**, 361 (1989).
- [38] W. Loinaz and R. S. Willey, Monte carlo simulation calculation of the critical coupling constant for two-dimensional continuum ϕ^4 theory, *Phys. Rev. D* **58**, 076003 (1998).
- [39] R. C. Brower and P. Tamayo, Embedded Dynamics for ϕ^4 Theory, *Phys. Rev. Lett.* **62**, 1087 (1989).
- [40] R. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1982).
- [41] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, How transferable are features in deep neural networks? in *Advances in Neural Information Processing Systems 27*, edited by Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger (Curran Associates, New York, 2014), pp. 3320–3328.
- [42] D. Schaich and W. Loinaz, Improved lattice measurement of the critical coupling in ϕ_2^4 theory, *Phys. Rev. D* **79**, 056008 (2009).
- [43] H. M. Yau and N. Su, On the generalizability of artificial neural networks in spin models, [arXiv:2006.15021](https://arxiv.org/abs/2006.15021).
- [44] T. Mendes-Santos, X. Turkeshi, M. Dalmonte, and A. Rodriguez, Unsupervised learning universal critical behavior via the intrinsic dimension, [arXiv:2006.12953](https://arxiv.org/abs/2006.12953).
- [45] A. Milchev, D. W. Heermann, and K. Binder, Finite-size scaling analysis of the ϕ^4 field theory on the square lattice, *J. Stat. Phys.* **44**, 749 (1986).
- [46] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).