Super-resolving the Ising model with convolutional neural networks

Stavros Efthymiou, Matthew J. S. Beach, and Roger G. Melko
Perimeter Institute for Theoretical Physics, Waterloo, Ontario N2L 2Y5, Canada and
Department of Physics and Astronomy, University of Waterloo, Waterloo N2L 3G1, Canada

(Dated: July 2, 2019)

Machine learning is becoming widely used in condensed matter physics. Inspired by the concept of image super-resolution, we propose a method to increase the size of lattice spin configurations using deep convolutional neural networks. Through supervised learning on Monte Carlo (MC) generated spin configurations, we train networks that invert real-space renormalization decimations. We demonstrate that super-resolution can reproduce thermodynamic observables that agree with MC calculations for the one and two-dimensional Ising model at various temperatures. We find that it is possible to predict thermodynamic quantities for lattice sizes larger than those used in training by extrapolating the parameters of the network. We use this method to extrapolate the exponents of the 2D Ising critical point towards the thermodynamic limit, which results in good agreement with theory.

I. INTRODUCTION

A primary challenge in the field of many-body physics is the efficient computational simulation of systems with a large number of particles or lattice sites. Such simulations are crucial for the investigation of strongly-interacting systems, the discovery of exotic phases of matter, and the design of new materials and devices. Recently, it has been proposed to treat many-body physics as a data-driven problem, where the large dimensionality of data motivates the exploration of machine learning algorithms.[1–3]

In condensed matter systems, neural networks were first used as supervised classifiers that distinguish phases and identify phase transitions, even in unconventional cases when there is no underlying order parameter.[1,4–6] As unsupervised applications, generative models have been designed that can capture thermal distributions.[7–9] In the quantum case, neural networks are actively explored as representations for many-body wavefunctions. They have been used as a variational ansatz[2] or for quantum state tomography.[10–12]

Early connections between statistical physics and machine learning drew parallels to the renormalization group[13,14] (RG), a canonical paradigm in physics which involves iteration through a series of coarse-graining and rescaling procedures. The mathematical similarity of the RG procedure to the processing of information in multilayer neural networks has driven interest in exploring the theoretical underpinnings of deep learning.[15,16] Conversely, relations between the RG and machine learning have proven useful for physics itself where neural networks have been proposed as generative models that assist RG procedures,[17] or for identifying relevant degrees of freedom to decimate.[18,19] Direct applications of neural networks on physical configurations can produce RG or inverse RG flows.[19,20]

Outside of physics, an area of expanding application for machine learning is image super-resolution, where the goal is to increase the number of pixels in an image while (subjectively) maintaining the perceivable quality.[21–23] Remarkable progress has been made with convolutional neural networks (CNNs), which can be used to reconstruct high-resolution images to photo-realistic quality.[22] In this paper, we investigate whether super-resolution methods may be useful in condensed matter and statistical physics by allowing one to produce lattice configurations of larger sizes directly from those obtained for smaller systems. For concreteness, we focus on the classical Ising model in one and two dimensions. Our method takes a configuration of Ising spins on a lattice, and subjects it to a majority rule block spin RG procedure.[24] A CNN is then trained to invert this transformation by being exposed to both the higher and lower resolution images. Since some information is necessarily lost in the RG step, the network output is interpreted as a probabilistic image, which is sampled to produce the super-resolved images.

Unlike previous works that explore analogies between deep learning and the RG,[16,20] here we force a relation through supervised training. We consider real-space decimation of the one-dimensional (1D) classical Ising model, along with the majority-rule block spin procedure for the two-dimensional (2D) model.[13] We give numerical evidence that the trained super-resolution network performs a probabilistic inverse of the RG transformation, and can reproduce thermodynamic quantities on a larger lattice size starting only from a smaller one. In addition, we propose a way to extrapolate the weights of a trained CNN to apply it to sizes larger than available in the training data. Using this idea iteratively, we acquire configurations of increasing size that we then use to estimate the critical exponents of the 2D Ising universality class, obtaining agreement with known theoretical results.

II. SUPER-RESOLUTION AND RG

In this section, we discuss the ensemble formed by super-resolved images generated from a collection of physical configurations or microstates. We also analyze the relevant similarities between super-resolution and the

RG. In particular, we explore how super-resolution is explicitly related to the real-space RG procedure.

Super-resolution is defined as a mapping $\mathcal{SR} : \mathbb{Z}_2^{L\times L} \to \mathbb{Z}_2^{fL\times fL}$ from a low-dimensional space of $L \times L$ images to a high-dimensional space of $(fL) \times (fL)$ images where $f > 1$ is the upscaling factor. We use $\mathbb{Z}_2$ to denote a binary variable, with $\mathbb{Z}_2^{L\times L}$ denoting an $L \times L$ matrix of binary values. The objective of image super-resolution in computer vision is to achieve a high perceived quality on the super-resolved image. This is generally a subjective criterion. To give a more quantitative definition, quantities like peak signal-to-noise ratio (SNR) or structural similarity (SIM) have been used.[21,22] However, even such quantities might not be a reliable estimator of the actual quality as perceived by a human.[22]

In contrast, statistical physics provides a well-defined objective for super-resolution of physical systems since super-resolved configurations should follow a specific statistical ensemble. Basic thermodynamic quantities like the magnetization or energy serve as an indicator of whether a super-resolved image is consistent with this ensemble. Figure 1 shows a configuration of the 2D Ising model at critically with some possible super-resolutions of the image.

Consider the real-space course-graining of the 2D Ising model according to the majority rule.[13] A $(2L) \times (2L)$ lattice is divided into 2×2 blocks where each block is transformed to a spin with the same state as the majority of spins in the block. If the total sign is zero, we take the sign of the upper left spin to make the procedure deterministic, instead of the more common probabilistic approach of taking a random sign.[25]

Let $\boldsymbol{\sigma} \in \mathbb{Z}_2^{(2L)\times(2L)}$ be an Ising configuration of N spins that follows the Boltzmann distribution

$$P_K(\boldsymbol{\sigma}) = \frac{1}{Z} e^{-H(\boldsymbol{\sigma})/T} = \frac{1}{Z} e^{K \sum_{\langle ij \rangle} \sigma_i \sigma_j}, \qquad (1)$$

where $K \equiv 1/T$ and $Z(K) = \sum_{\{\boldsymbol{\sigma}\}} e^{-H(\boldsymbol{\sigma})/T}$ is the partition function. We take periodic boundary conditions (PBC) and include only nearest-neighbor interactions throughout this paper.

A low-resolution configuration, $\mathbf{s} \equiv \mathcal{MR}(\boldsymbol{\sigma}) \in \mathbb{Z}_2^{L\times L}$, is obtained upon applying the deterministic majority rule. Such a configuration follows the marginalized distribution:

$$\tilde{P}_K(\mathbf{s}) = \sum_{\{\boldsymbol{\sigma}\}} k(\mathbf{s}, \boldsymbol{\sigma}) P_K(\boldsymbol{\sigma}), \qquad (2)$$

where k is the kernel of the transformation.[26] In this way, the distribution of low-resolution configurations, $\mathbf{s}$, related to the distribution of high-resolutions $\boldsymbol{\sigma}$'s.

Evidently, any super-resolution procedure must satisfy

$$\mathcal{MR}(\mathcal{SR}(\mathbf{s})) = \mathbf{s}\,. \qquad (3)$$

A stronger requirement, and the main challenge, is to discover a map such that $\mathcal{SR}(\mathcal{MR}(\boldsymbol{\sigma}))$ obeys the correct Boltzmann distribution. We emphasize that $\mathcal{SR}(\mathcal{MR}(\boldsymbol{\sigma}))$ is not necessarily equal to $\boldsymbol{\sigma}$ since only the distributions need to match, not each individual configuration. Furthermore, in our majority rule decimation, we have not rescaled the Hamiltonian couplings as required in the conventional definition of a complete RG step. It may be possible to learn the rescaling with a neural network; however, in this paper we simply numerically fit the couplings as needed for a consistent rescaling, as discussed in detail below.

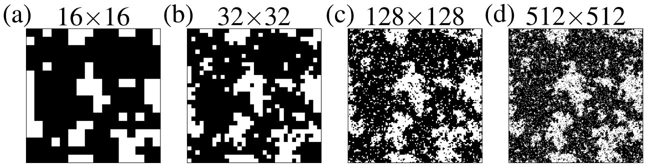

FIG. 1. Critical configurations obtained using the weight extrapolation idea presented in Section II B. We show the original Monte Carlo configuration in (a) and the results after (b) one, (c) three and (d) five consecutive super-resolutions.

A. Network architecture

We now attempt to invert the majority rule procedure using a supervised learning approach. The unknown super-resolution mapping $\mathcal{SR}$ is parametrized with a deep convolutional neural network (CNN), $\mathcal{F}_{\mathcal{W}}$ where $\mathcal{W} = \{W_{ij}, b_i\}$ are weights and biases. The parameters $\mathcal{W}$ are tuned by minimizing a loss function on a dataset of inputs $\mathbf{s}_i \in \mathbb{Z}_2^{L\times L}$ and targets $\boldsymbol{\sigma}_i \in \mathbb{Z}_2^{(2L)\times(2L)}$ with $i \in \{1, 2, \ldots, n\}$. In contrast to typical supervised learning applications (e.g. handwritten digit recognition) the dimensionality of the output is larger than the input.

The first layer of our CNN is an upsampling layer that increases the resolution from $L \times L$ to $(2L) \times (2L)$ by transforming each up/down spin to a block of four up/down spins. The convolutions that follow add the required noise, in a way similar to the Monte Carlo sweep procedure in Ref. [24]. In each layer, we use PBC padding to preserve the $(2L) \times (2L)$ lattice size while remaining consistent with the boundary conditions of the model.

A sigmoid activation in the last layer gives a continuous output $\mathbf{p}$ with each element $p_k \in [0, 1]$. To transform this into the discrete Ising configuration, we treat the output neurons as the probability that the corresponding spin is up. Hence, the discrete configuration is obtained by sampling from the uniform distribution at each lattice site.

To find the optimal parameters $\mathcal{W}$, we minimize a loss function on the training dataset. The loss quantifies the distance between predicted output $\mathcal{F}_{\mathcal{W}}(\mathbf{s}_i)$ and the original high-resolution $\boldsymbol{\sigma}_i$. Minimization is done with backpropagation[27] which involves calculations of gradients, and thus cannot be done using the final sampled (discrete) output. To be consistent with the interpretation of the continuous outputs as probability we use the cross-entropy loss function;

$$\mathbb{L}(\{\boldsymbol{\sigma}_i\},\{\mathbf{s}_i\}) = \sum_{i=1}^{n} \left[\boldsymbol{\sigma}_i \cdot \ln \mathbf{p}_i + (1-\boldsymbol{\sigma}_i)\cdot \ln\left(1-\mathbf{p}_i\right)\right], \tag{4}$$

where $i \in \{1, 2, \ldots, n\}$ and $\cdot$ denotes the element-wise dot product between matrices.

We are free to add additional terms to Eq. (4) to assist training. As we will see, it is sometimes beneficial to introduce a term $\gamma|E(\boldsymbol{\sigma}_i) - E(\mathbf{p}_i)|^2$ to enforce that the energy be similar between a high-resolution $\boldsymbol{\sigma}$ and the corresponding super-resolved configuration.

B. Extrapolation to larger lattice sizes

Using the method discussed above, we define a super-resolution mapping in 2D as $\mathcal{SR} : \mathbb{Z}_2^{L\times L} \to \mathbb{Z}_2^{(2L)\times(2L)}$. In order to train the network $\mathcal{F}_\mathcal{W}$ that applies this transformation, we need access to $(2L)\times(2L)$ configurations which we obtain with Monte Carlo simulations. In this sense, the method does not allow us to access sizes larger than the ones we have already simulated.

From a practical perspective, it would be useful if super-resolution could be used to access sizes that cannot be obtained by other means. We propose a simple method to do precisely this by exploiting the weight sharing property of convolutions. Namely, weight matrix dimensions on a convolutional layer depend only on the kernel size and channel number, which are independent of the input and output size. In order to apply the convolution to a larger image, we only have to "slide" the same kernel over a larger surface. The first upsampling layer does not contain any weights and can be trivially applied to any size. Therefore, in principle, we can apply our trained network to double any input size. Using this property repeatedly, taking the output of each super-resolution as the new input, we can access arbitrary large sizes by creating a chain of increasing sizes ($\mathbb{Z}_2^{16\times 16} \to \mathbb{Z}_2^{32\times 32} \to \mathbb{Z}_2^{64\times 64} \to \ldots$). This procedure works in any dimension, given the correct filters.

III. ONE-DIMENSIONAL ISING MODEL

In this section, we show the results of the super-resolution scheme applied to the 1D Ising model.

We use a dataset consisting of temperatures ranging from $T = 0.01$ to $T = 3.5$ with $N = 32$ spins. At each temperature, we create training and testing sets consisting of $n = 10^4$ configurations generated via standard Monte Carlo. Instead of the majority rule, we use real-space block-spin decimation to obtain a $N = 16$ chain.[14] For each temperature, we implement a different network and optimize using Adam,[28] with a batch size of 10^3. Instead of training for a specific amount of epochs, we cease training when the validation loss stops improving using early stopping.

We find that the network achieves reasonable accuracy at each temperature when evaluated in the test set (Fig. 8 in Appendix B). This shows that the network performs an approximate inverse of the block-spin decimation. We note that this inversion is demonstrated only at the level of thermodynamic observables and not the whole statistical ensemble. We further show that super-resolution can generate new data for sizes larger than the ones used in training, by using the extrapolation technique of Section II B.

Until now we have neglected the rescaling part of an RG step, but it will be essential to obtaining larger configurations. Conveniently, in 1D we can exactly calculate the marginalization of Eq. (2). The Hamiltonian is self-similar under the RG decimation with the rescaled couplings

$$\tilde{K} = \frac{1}{2}\ln\cosh 2K\,. \tag{5}$$

Under successive RG steps, the temperature T of a configuration flows towards infinity.

Because the network is trained on pairs $(\mathbf{s}_i, \boldsymbol{\sigma}_i)$ without this temperature adjustment, it is necessary to correct this manually when making predictions on inputs $\boldsymbol{\sigma}$ that follows the Boltzmann distribution of Eq. (1), instead of the marginalized distribution of Eq. (2). Therefore, given an input $\boldsymbol{\sigma}$, at temperature $\tilde{T}$, the network produces configurations $\mathcal{F}_\mathcal{W}(\boldsymbol{\sigma})$ at a new temperature T obtained by inverting Eq. (5).

To validate our extrapolation proposal, we compare the magnetization and energy of super-resolved configurations $\mathcal{F}_\mathcal{W}(\boldsymbol{\sigma}) \in \mathbb{Z}_2^{64}$, with configurations $\boldsymbol{\sigma}^{(64)} \in \mathbb{Z}_2^{64}$ generated via Monte Carlo. Fig. 2 shows the results of the network after adjusting the temperature with Eq. (5). We stress that the network was trained on $\mathbb{Z}_2^{16} \to \mathbb{Z}_2^{32}$ data and yet predicts $N = 64$ data by exploiting the extrapolation idea from Section II B.

We see that, even though the network has never seen $N = 64$ data before, it can approximate the magnetization M and energy E within a few percent. We expect that by increasing the extrapolation to larger sizes, any error in the original data or imperfections in the network will propagate. However, we suggest that for a single extrapolation, the network does remarkably well. Even multiple extrapolations may serve as a good starting point for large-scale simulations, e.g. possibly shortening equilibration time in a Monte Carlo procedure.

As another test of the super-resolving network, we repeat extrapolation up to $N = 512$ spins and we calculate the two-point function, $G_N(j) = \langle \sigma_1 \sigma_{1+j}\rangle$, for configurations from each generated size. In 1D, the exact value for this quantity is:[29]

$$G_N(j;K) = \frac{\tanh^j K + \tanh^{N-j} K}{1+\tanh^N K}\,. \tag{6}$$

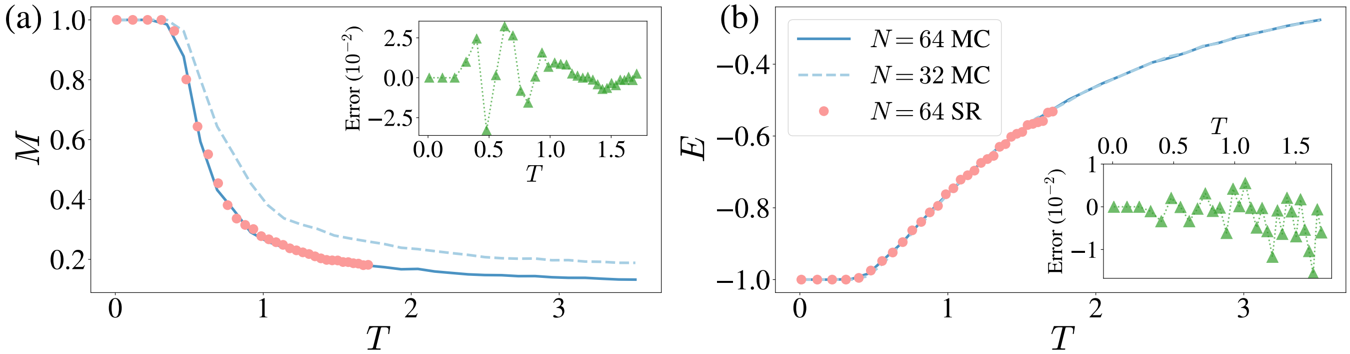

FIG. 2. (a) Absolute magnetization and (b) energy per spin for the 1D Ising model. We denote Monte Carlo results at low ($N = 32$) and high ($N = 64$) resolution with MC. The super-resolution (SR) results were obtained by using the low resolution MC as input to an extrapolated $\mathbb{Z}_2^{16} \to \mathbb{Z}_2^{32}$ network. SR temperatures are adjusted according to the inverse of Eq. (5). This shrinks the temperature range as the inverse RG transformation flows towards $T = 0$. Inset plots correspond to the error between SR predictions and MC results.

In Fig. 3 we plot the two-point function for the different sizes and $j = N^{0.8}/5$. The lines correspond to Eq. (6), while the marked points correspond to those computed via super-resolved configurations. We note that the choice $j = N^{0.8}/5$ does not have a particular physical significance as it is possible to obtain similar accuracy for different choices of j. The same effects are also observed in other estimators. Generally, errors are expected to increase with each super-resolution step. However, in the current case, the temperature flows towards zero under the inverse Eq. (5), so the extrapolation scheme remains accurate even after multiple super-resolutions.

We have demonstrated numerically with two different methods that our super-resolution mapping can successfully capture thermodynamic quantities, as an approximate inverse RG transformation. The network parameter extrapolation is particularly effective in 1D where we know exactly how to rescale from Eq. (5).

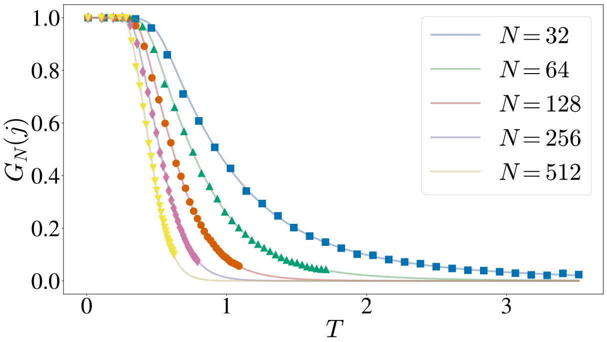

FIG. 3. Two-point function of the 1D Ising model with $j = N^{0.8}/5$. Solid lines correspond to Eq. (6) and marked points to the super-resolution prediction. We use MC data for $N = 32$, while all other sizes are obtained from consecutive super-resolutions, extrapolating the $\mathbb{Z}_2^{16} \to \mathbb{Z}_2^{32}$ network. Temperatures are adjusted by repeatedly applying the inverse of Eq. (5).

IV. TWO-DIMENSIONAL ISING MODEL

Following the 1D case, we proceed by training a 2D $\mathbb{Z}_2^{8\times 8} \to \mathbb{Z}_2^{16\times 16}$ network using Monte Carlo generated datasets. The majority rule is now used instead of the simple decimation. We again find success on the testing set (Appendix B) and seek to explore extrapolation to larger sizes.

A challenge in 2D is that the marginalization of Eq. (2) cannot be done analytically and hence we cannot simply rescale using Eq. (5) as we did in 1D. The 2D model is not self-similar under the block-spin RG transformation, as the Hamiltonian that corresponds to the decimated distribution contains interactions beyond nearest-neighbors.[30]

To obtain a 2D analogy of Fig. 2, we perform the temperature correction numerically. During training, the temperature T assigned to the down-sampled (DS) configurations is not the same as the temperature $\tilde{T}$ of those configurations if they occurred from MC samples. Likewise, the SR configurations at T differ from large-lattice MC data. The function $f : T \to \tilde{T}$ which takes the DS temperature to the MC temperature, also takes the larger-lattice MC temperature to the SR one. While the RG rescaling f pushes the temperature to the stable fixed points ($T = 0, \infty$), the SR rescaling f^{-1} pushes the temperature towards critically (T_c). In order to numerically find the transformation $f^{-1} : \tilde{T} \to T$ we compare observables calculated on a 8×8 MC configuration with those calculated on a 8×8 decimated one. We require that the corresponding curves collapse when the transformation is applied on the MC curve.

We note that this has no direct physical interpretation since the 2D Ising Hamiltonian is not self-similar to the block spin RG transformation and therefore, temperature

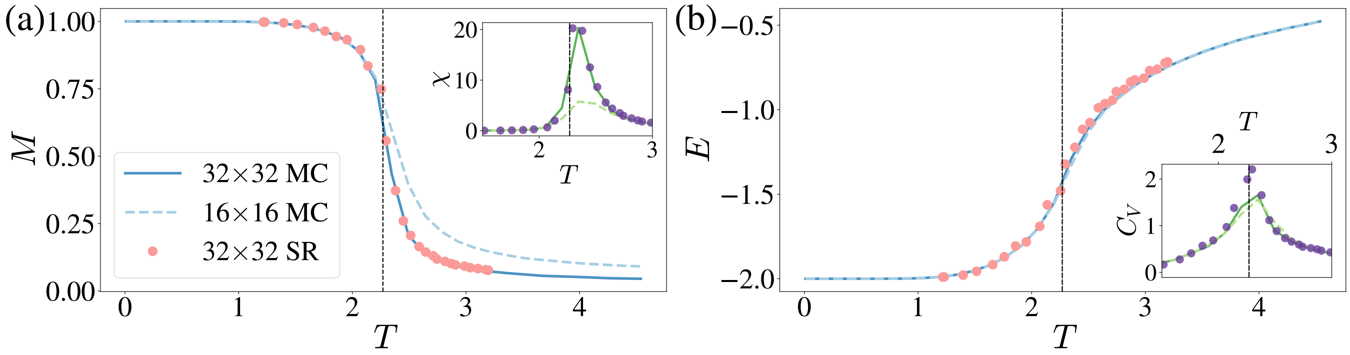

FIG. 4. (a) Magnetization (with susceptibility) and (b) energy (with specific heat) for the 2D Ising model. MC denotes Monte Carlo results while SR is obtained by super-resolving the 16 × 16 MC configurations using the extrapolation of the $\mathbb{Z}_2^{8\times 8} \to \mathbb{Z}_2^{16\times 16}$ network. SR temperatures were rescaled using the numerical transformation, which shrinks temperature range towards criticality.

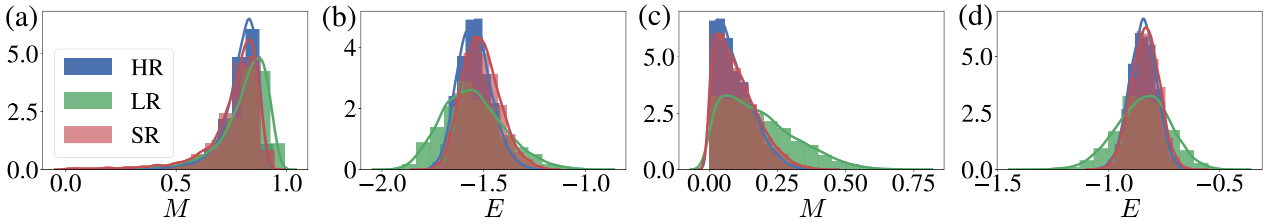

FIG. 5. Histograms of the magnetization and energy distributions at (a, b) $T = 2.2010$ and (c, d) $T = 2.9313$. Thermodynamic observables are binned into 20 bins, and the smooth line shows a kernel density estimator from the Seaborn module. HR (high-resolution) denotes the 32 × 32 Monte Carlo configurations, LR (low-resolution) the 16 × 16 Monte Carlo, and SR the results by extrapolating the $\mathbb{Z}_2^{8\times 8} \to \mathbb{Z}_2^{16\times 16}$ network.

alone is not sufficient to describe the coupling space of the RG configuration. Here, we use this procedure only to demonstrate that our results in Fig. 4 are consistent with the inverse-RG nature of super-resolution. This follows from the fact that the same transformation that makes MC observables collapse to DS observables, also makes the SR observables collapse to MC predictions.

We proceed by extrapolating the parameters of the trained $\mathbb{Z}_2^{8\times 8} \to \mathbb{Z}_2^{16\times 16}$ network and using it to super-resolve 16×16 MC configurations to 32×32. We present the results for magnetization and energy in Fig 4, where susceptibility and specific heat are also shown. Temperatures for the SR observables are rescaled according to the $\tilde{T} \to T$ transformation as described above. As in 1D, we see that the rescaling makes predicted SR observables match the MC results, indicating again that the network performs an approximate inversion of the RG transformation.

Low temperatures are trivial because all spins are usually polarized in one direction and it is not required to add any noise during the super-resolution. At high temperatures, the noise is generally random and difficult to learn. In contrast to 1D, we add a regularization term in the loss function, which compares the energy of the super-resolved configuration to that of the original one. This does not use any more information than already present in the training data, but results in better convergence of the network for high temperatures.

To corroborate our findings, we calculate the probability distributions of magnetization and energy. We present them in the form of histograms in Fig. 5, for $T \simeq T_c$ and $T > T_c$. As in Fig. 4, the 32×32 super-resolution (SR) results are obtained by extrapolating the $\mathbb{Z}_2^{8\times 8} \to \mathbb{Z}_2^{16\times 16}$ network and using it on the low-resolution (LR) 16×16 Monte Carlo configurations. Input temperatures are adjusted to agree with the approximate inverse RG transformation. The histograms are created by calculating magnetization and energy for each one of the 10,000 configurations in the test set and binning the resulting values into 20 bins. To get the approximate probability density function, that is the smooth line fitted in the histograms of Fig. 10, we use a kernel density estimator[31] provided by the Seaborn module.[32] The estimator generates a Gaussian distribution centered in each dataset point and approximates the target probability density as the sum of these distributions normalized. We expect a reliable approximation because the distributions are localized around a point and do not have heavy tails. We see that super-resolution captures not only the average values of magnetization and energy, but their entire probability distribution.

A. Critical Exponents

An interesting application of super-resolution in the 2D case is the calculation of critical exponents. In principle, we can circumvent the problem of self-similarity if we focus on the fixed point of the RG transformation. Here we crudely approximate[24] the fixed point with the nearest-neighbor Hamiltonian precisely at the critical temperature. We take 10^5 samples of 16×16 Monte Carlo configurations and we repeatedly extrapolate to reach sizes up to 512×512. Examples of the acquired configurations are shown in Fig. 1. We stress that Monte Carlo simulation is required only on the smallest size (in our case 16×16) as all larger sizes are obtained by extrapolating the $\mathbb{Z}_2^{8\times 8} \rightarrow \mathbb{Z}_2^{16\times 16}$ network.

We use the predicted configurations to calculate the critical exponents for the 2D Ising universality class using the finite-size scaling hypothesis.[33] According to this, exactly at criticality $\chi \propto L^{\gamma/\nu}$, where γ and ν are the susceptibility and correlation length critical exponents respectively. We can estimate the γ/ν ratio from the slope in a log-log (L, χ) plot. Similarly, the two-point function vanishes algebraically $G(r) \sim 1/r^{d-2+\eta}$ at criticality, allowing us to estimate the anomalous dimension η from the slope of a log-log $(r, G(r))$ plot. The magnetization exponent β can be calculated similarly.

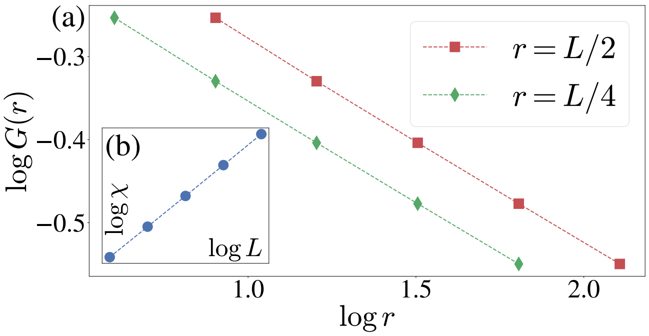

FIG. 6. Scaling of the (a) two-point function and (b) susceptibility at criticality. The smallest size is calculated with Monte Carlo and the rest with repeated super-resolutions. Errors are typically around $\sim 10^{-3}$ and too small to show in this figure.

We calculate the values of these quantities at every level of the repeated inverse RG procedure and give the corresponding log-log plots in Fig. 6. The two-point function is calculated using two different values of the corresponding distance $r = L/4$ and $r = L/2$. The plots confirm the expected linear behavior. To get an estimation of our method's error in the critical exponents, we repeat the training and critical exponent calculation 60 times. In Table I, we give predicted values and percentage error. In the last column, we give the results from Monte Carlo simulations. These exponents were calculated using 30,000 configurations at each sizes for sizes 4×4, 8×8, 12×12 and 16×16. We emphasize that this is a fair comparison since the network *only* had access to 16×16 (and decimated) configurations both in the training and calculation procedure.

Exponent	Super-resolution	Error	Finite-size scaling error
β	0.1234 ± 0.006	1.3%	1.3%
γ	1.7544 ± 0.01	0.25%	2.3%
η_1	0.2460 ± 0.01	1.6%	2.3%
η_2	0.2459 ± 0.01	1.6%	1.1%

TABLE I. Critical exponents of the 2D Ising universality class. We give the mean and standard error of 60 independent repetitions of training and critical exponent calculation from 16×16 to 128×128. In the third column, we compare our results to theoretical values, while in the last column we give results obtained from MC configurations with size up to 16×16.

We note that the linearity of scaling can be confirmed from the correlation coefficient of the regression which differs from unity less than 10^{-4} in all cases. The y-intercept in the χ fit agrees with the Monte Carlo calculation in Fig. 14 of Ref. [33]. The most important result is that we achieve less than 2% error from the theoretical value of the exponents in all cases.

V. SUMMARY

We have investigated whether super-resolution techniques can be used to successfully increase the size of physical configurations sampled from the 1D and 2D Ising Hamiltonian. Inspired by recent applications of deep learning, we used a convolutional neural network for this task. We performed supervised training with a set of Monte Carlo configurations as output, and their corresponding RG decimated counterparts as input. Therefore, the network was essentially trained to double the size of configurations by performing a transformation approximately equivalent to an inverse RG step.

Despite the challenge in rigorously defining the inverse RG transformation due to the loss of information during the decimation, we found that our super-resolution scheme can accurately capture some thermodynamic observables over a wide range of temperatures. We further proposed a method to extrapolate the trained weights/biases and used them to access arbitrary lattice sizes larger than those used for training. We found that the extrapolation worked well for both 1D and 2D systems, and we were able to compute critical exponents in the latter case which show agreement with analytical results to within 2%.

Such techniques may be beneficial to the large-scale simulation of complex physical systems, where our extrapolation may provide approximate initial configurations for further optimization procedures. One obvious application would be to try super-resolved configurations to shorten equilibration times in Monte Carlo methods. Decimation and super-resolution could also be used to attempt to define non-local updating procedures in models that suffer from long autocorrelation times.

It would be interesting to explore more sophisticated RG procedures beyond the real-space Ising model. Ideas analogous to the weight extrapolation might allow one to simulate models on lattice sizes that are inaccessible by other means. A quantum generalization could provide a way to generate approximate configurations that are inaccessible to conventional simulation techniques. Conversely, further exploration of how the neural networks identify the loss of information between original and decimated configurations might contribute to a fundamental understanding of information propagation in deep learning techniques.

ACKNOWLEDGMENTS

The authors would like to thank J. Carrasquilla, A. Golubeva, L. E. Hayward Sierens, B. Kulchytskyy, P. Ponte, I. Tamblyn, and G. Torlai, for many useful discussions. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chair program, and the Perimeter Institute for Theoretical Physics. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET) and Compute/Calcul Canada. We also gratefully acknowledge the support of NVIDIA Corporation with the donation of the Titan Xp GPU used for this research. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Research & Innovation.

[1] J. Carrasquilla and R. G. Melko, Nature Physics **13**, 431 EP (2017).

[2] G. Carleo and M. Troyer, Science **355**, 602 (2017).

[3] Y. Zhang, A. Mesaros, K. Fujita, S. D. Edkins, M. H. Hamidian, K. Ch'ng, H. Eisaki, S. Uchida, J. C. Séamus Davis, E. Khatami, and E.-A. Kim, (2018), arXiv:1808.00479.

[4] L. Wang, Phys. Rev. B **94**, 195105 (2016).

[5] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Phys. Rev. X **7**, 031038 (2017).

[6] M. J. S. Beach, A. Golubeva, and R. G. Melko, Phys. Rev. B **97**, 045207 (2018).

[7] G. Torlai and R. G. Melko, Phys. Rev. B **94**, 165134 (2016).

[8] Z. Liu, S. P. Rodrigues, and W. Cai, (2017), arXiv:1710.04987.

[9] S. J. Wetzel, Phys. Rev. E **96**, 022140 (2017).

[10] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nature Physics **14**, 447 (2018).

[11] G. Torlai and R. G. Melko, Phys. Rev. Lett. **120**, 240503 (2018).

[12] *QuCumber* (2018) github.com/PIQuIL/QuCumber.

[13] J. Cardy, *Scaling and Renormalization in Statistical Physics* (Cambridge University Press, 1996).

[14] L. Kadanoff, *Statistical Physics: Statics, Dynamics and Renormalization* (World Scientific Publishing Company, 2000).

[15] C. Beny, (2013), arXiv:1301.3124v4.

[16] P. Mehta and D. J. Schwab, (2014), arXiv:1410.3831.

[17] S.-H. Li and L. Wang, (2018), arXiv:1802.02840v3.

[18] M. Koch-Janusz and Z. Ringel, Nature Physics **14**, 578 (2018).

[19] P. M. Lenggenhager, Z. Ringel, H. S. D., and M. Koch-Janusz, (2018), arXiv:1809.09632.

[20] S. Iso, S. Shiba, and S. Yokoo, Phys. Rev. E **97**, 053304 (2018).

[21] C. Dong, C. C. Loy, K. He, and X. Tang, IEEE Transactions on Pattern Analysis and Machine Intelligence **38**, 295 (2016).

[22] C. Ledig, L. Theis, F. Huszr, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2017) pp. 105–114.

[23] C. Dong, C. C. Loy, and X. Tang, in *Computer Vision – ECCV 2016*, edited by B. Leibe, J. Matas, N. Sebe, and M. Welling (Springer International Publishing, Cham, 2016) pp. 391–407.

[24] D. Ron, R. H. Swendsen, and A. Brandt, Phys. Rev. Lett. **89**, 275701 (2002).

[25] R. H. Swendsen, Phys. Rev. Lett. **42**, 859 (1979).

[26] M. Kardar, *Statistical Physics of Fields* (Cambridge University Press, 2007).

[27] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Nature **323**, 533 EP (1986).

[28] D. P. Kingma and J. Ba, CoRR **abs/1412.6980** (2014), arXiv:1412.6980.

[29] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, London, 1989).

[30] R. H. Swendsen, Phys. Rev. B **30**, 3866 (1984).

[31] E. Parzen, Ann. Math. Statist. **33**, 1065 (1962).

[32] M. Waskom, *Seaborn: Statistical Data Visualization* (2012).

[33] A. W. Sandvik, AIP Conference Proceedings **1297** (2010).

Appendix A: Importance of sampling

As mentioned in Section II, the continuous output after the last sigmoid layer is interpreted as the probability that the spin in the corresponding site is up. Therefore, the super-resolved configuration is obtained by sampling this continuous output. The $\mathcal{F}_{\mathcal{W}} : \mathbb{Z}_2^{L\times L} \to \mathbb{Z}_2^{2L\times 2L}$ mapping consists of the convolutional network and the sampling procedure.

This sampling procedure makes the $\mathcal{SR}$ mapping non-deterministic. We believe that sampling is crucial for the method to work. Here we give some numerical evidence to corroborate this statement.

The loss of information in the majority rule RG is associated with the different types of blocks that give the same decimated spin. For example, consider a block with 4 up - 0 down spins and one with 3-1. Both would lead to an up spin in the decimated configuration. In order to capture the correct thermodynamics, the inverse RG procedure should give these different block types with the correct proportion at each temperature. To investigate whether this happens, we observe that the type of each block can be uniquely defined by the sum of spins contained in the block. In the $\mathbb{Z}_2 = \{0, 1\}$ convention, this sum is the number of up spins and goes from 0 to 4.

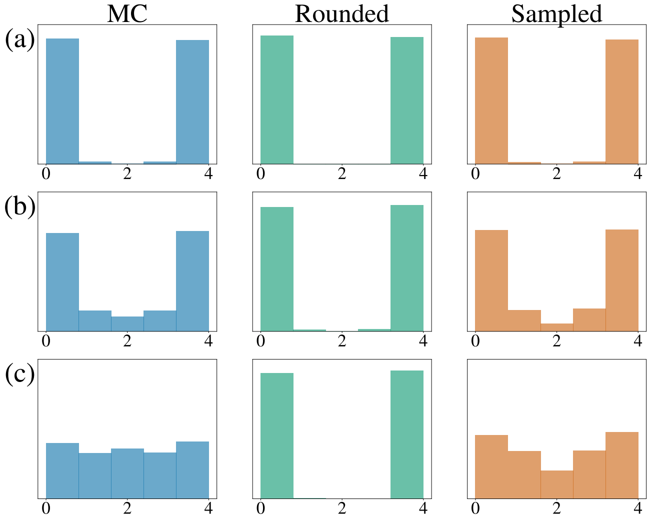

FIG. 7. Histograms of the different 2×2 block sums at three different temperatures (a) $T = 1.4706$ (low), (b) $T = 2.2010$ (critical) and (c) $T = 2.9313$ (high). The first column corresponds to MC configurations, while the second and third to rounded and uniformly sampled network output respectively.

In Fig. 7 we give the number of appearances of each block sum in the original MC configurations and different interpretations of the network's output (rounding or sampling). The height of each bar is calculated by summing the appearances of each block sum over each configuration. At low temperatures, most configurations are fully polarized with the value 0 (all down) and 4 (all up). In this case, we do not have information loss during RG and there is no significant difference between rounding and sampling the output. At temperatures near and above criticality, non-fully polarized blocks start to appear, increasing the appearance of intermediate sums (1 to 3). Rounding fails to capture the non-fully polarized blocks, making imperative the use of sampling. Even sampling cannot accurately capture the blocks with 2-2 spins, indicating an inaccuracy in our method which could be improved in further work.

Appendix B: Testing Data

In Sections III and IV we demonstrated that super-resolution can capture the thermodynamics of sizes larger than the ones used in training, simply by extrapolating the network parameters. Before this step, it is important to directly evaluate the network's performance on the objective that it was trained on, namely the super-resolution of down-sampled configurations. We present this evaluation here.

In Fig. 8 we test the 1D $\mathbb{Z}_2^{16} \to \mathbb{Z}_2^{32}$ network on super-resolving down-sampled (DS) $N = 16$ configurations and we show that it correctly captures the $N = 32$ MC results. Fig. 9 gives the same test for the 2D $\mathbb{Z}_2^{8\times 8} \to \mathbb{Z}_2^{16\times 16}$ network. As in the main text, 2D results are corroborated by the histograms of Fig. 10, which show that the whole magnetization and energy distributions predicted by super-resolution match the corresponding MC distributions. In each dimension, DS configurations are obtained by applying the respective decimation transformation on the large MC samples. DS configurations are then used as the network's input for trainin. Therefore, these results indicate directly that the network approximately inverts the down-sampling procedure, the caveat being the underrepresentation of 2-2 blocks. Note that, unlike the main text, in the current appendix we did not rescale the temperature of SR predictions. Rescaling is not needed because the input is decimated configurations that follow the marginalized distribution in Eq. (2).

As expected, the errors are typically lower for the direct testing case presented in the current Appendix, compared to the extrapolation results of Section III. Also, errors are generally larger at higher temperatures, where thermal noise becomes more random and more difficult to learn.

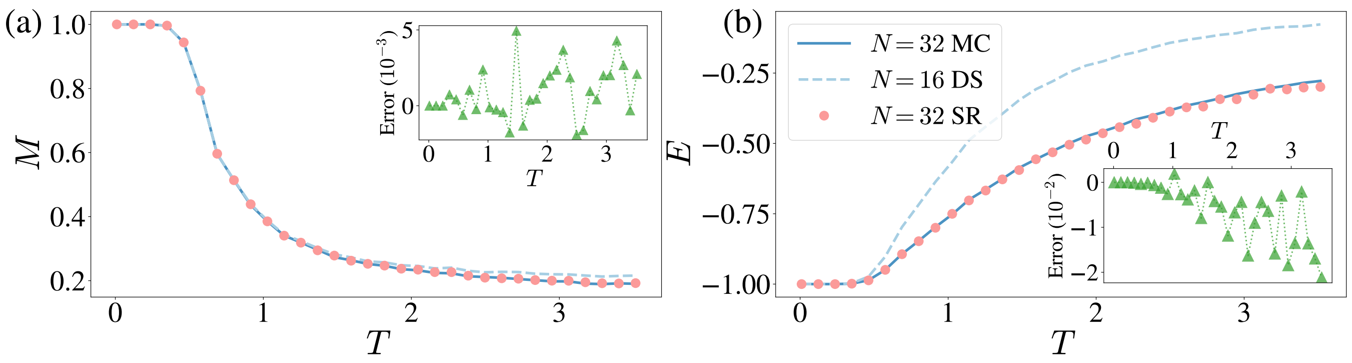

FIG. 8. (a) Magnetization and (b) energy of the 1D Ising model. The dashed line corresponds to observables computed with the down-sampled (DS) configurations used as the network's input.

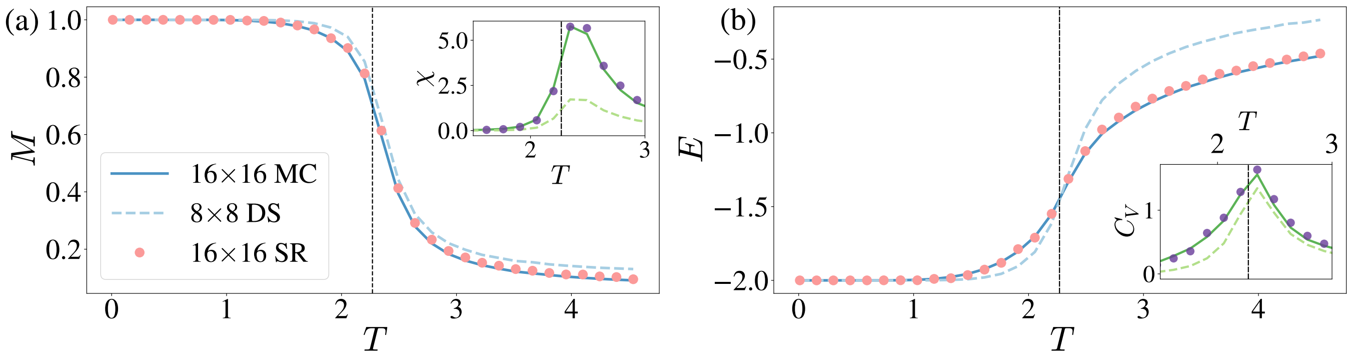

FIG. 9. (a) Magnetization (with susceptibility) and (b) energy (with specific heat) of the 2D Ising model.

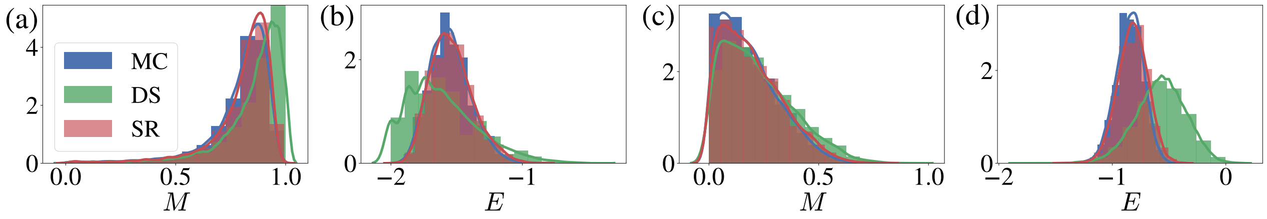

FIG. 10. Histograms of the magnetization and energy distributions at (a, b) $T = 2.2010$ and (c, d) $T = 2.9313$. Thermodynamic observables are binned into 20 bins, and the smooth line shows a kernel density estimator from the Seaborn module. MC denotes 16×16 Monte Carlo configurations, DS the 8×8 majority rule down-sampled configurations, and SR the results upon applying the $\mathbb{Z}_2^{8\times 8} \to \mathbb{Z}_2^{16\times 16}$ network on DS.