Fine-tuning Vision Transformers for the Prediction of State Variables in Ising Models

Onur Kara[*]
okara83@gmail.com

Arijit Sehanobish[†]
arijit.sehanobish1@gmail.com

Hector H. Corzo[‡]
hhcorzo@gmail.com

Abstract

Transformers are state-of-the-art deep learning models that are composed of stacked attention and point-wise, fully connected layers designed for handling sequential data. Transformers are not only ubiquitous throughout Natural Language Processing (NLP), but also have recently inspired a new wave of Computer Vision (CV) applications research. In this work, a Vision Transformer (ViT) is finetuned to predict the state variables of 2-dimensional Ising model simulations. Our experiments show that ViT outperforms state-of-the-art Convolutional Neural Networks (CNN) when using a small number of microstate images from the Ising model corresponding to various boundary conditions and temperatures. This work explores the possible of applications of ViT to other simulations and introduces interesting research directions on how attention maps can learn the underlying physics governing different phenomena.

1 Introduction

The Ising model is regarded as the simplest theoretical framework to describe ferromagnetism, and understand phase transitions [1]. The 2D Ising model is a mathematical model of atomic spins on a lattice which exhibits a phase transition that can be computed analytically [2]. The system undergoes a second order phase transition at the critical temperature T_c. In particular, when $T << T_c$, it undergoes spontaneous magnetization, and this phenomena characterizes ferromagnetism or the ordered state. The local interaction between spins is ultimately responsible for this behavior. For high temperatures, $T >> T_c$ the system is in the disordered or the paramagnetic state. In this case, there are no long-range correlations between the spins. Ferromagnetism is a collective phenomena which occurs when the spins of atoms in a lattice align such that the associated magnetic moments all point in the same direction. Analytical and numerical solutions for the Ising model have been extremely important in the understanding of phase transitions and ferromagnetism. However, in many cases, these solutions are extremely difficult to formulate and compute. Thus various machine learning (ML) methods have been used to understand these phase transitions.

Challenges in applying ML techniques to dynamical Ising models include the large number of interacting degrees of freedom and the so-called quenched disorder [3], the latter of which might result in a system evolving without temporal significance. Attempts to overcome these problems in Ising like systems using ML date back to [4, 5, 6]. Recently, the successes of deep learning and its applications to complex physical systems have reinvigorated interest in the field [7, 8, 9, 10, 11, 12]. All of the recent work has been focused on using CNN for understanding the Ising systems near the critical temperature [13, 14, 15, 16, 17, 18]. Several authors have also used CNNs in Generative Adversarial Networks (GAN) [19] and in Variational Autoencoders [20, 21] to generate images that simulate the system near the critical temperature. Given the success of transformers in CV, it is

[*] Hindsight Technology Solutions

[†] Work done as a postdoc at Yale University. Currently at Covera Health, NYC

[‡] Center for Chemical Computation and Theory, University of California, Merced, CA.

Fourth Workshop on Machine Learning and the Physical Sciences (NeurIPS 2021).

	FSbCR	SbCR	Cr	SpCR
Train	220	80	90	210
Validation	100	50	60	90
Test	150	50	70	130

Table 1: Data distribution: Number of images in each bin for each boundary condition.

natural to ask whether a ViT can match or outperform a CNN in understanding the phase transitions in an Ising model.

In this work, our contributions are the following: **(a)** we created a custom-made suite to generate high resolution Ising grid images for a large number of systems with different boundary conditions at various temperatures, **(b)** a ViT is fine-tuned on these images to predict the state variables corresponding to each of the simulation's experimental constraints, and finally, **(c)** to the best of our knowledge, this is the first time ViT has been used to understand classical statistical mechanical phenomenon in the Ising model and we show that ViT outperforms CNN based architectures like ResNet-18 and ResNet-50 when using a small number of labeled images.

2 Ising Model

The Hamiltonian of a system expresses the total energy of the system in question. Classically, the Hamiltonian is understood as the sum of the kinetic and potential energies. In the case of the two-state Ising Model, the standard Hamiltonian, $\mathcal{H}(\sigma)$, reads

$$\mathcal{H}(\sigma) = -\sum_{\langle i,j \rangle} J_{i,j}\sigma_i\sigma_j - B\sum_i \sigma_i \tag{1}$$

where $J_{i,j}$ is the interaction strength between the ith and jth spin sites, B is the external magnetic field, and $\sigma_i = \pm 1$ is the ith spin on the grid usually taken, as is the case here, to be an $m \times n$ lattice. The sum $\langle i,j \rangle$ is taken over all sites without double counting, and typically, the interaction $J_{i,j}$ is constant for all nearest neighbors across the lattice. In all our systems, $\mathcal{H}(\sigma)$ is defined according to the availability of spin sites in a restricted finite volume. Starting at T$_c$ for Ising Models, the order parameter of the second-order phase transitions increases continuously from zero. Second-order phase transitions are characterized by a high temperature phase with an average zero magnetization (disordered phase) and a low temperature phase with a non-zero average magnetization. This is very well demonstrated in the Metropolis-Hastings type Ising model when observing the continuous increase of the magnetization at a ferromagnetic-paramagnetic phase transition. When J is positive, the spins align parallel to one another, indicating ferromagnetic behavior. On the other hand, when coupling constant J is negative, nearest neighbor spins are in their lowest energy state by aligning themselves anti-parallel. In that case, the system is anti-ferromagnet. The properties of the anti-ferromagnetic 2D Ising model behaves almost identically to the ferromagnetic Ising model. However, in the lowest energy state $T \approx 0$ a checkerboard-like configuration represents the lowest energy state in contrast to the ferromagnetic case where energy is minimized with all spins in one direction or the other (fig 1).

3 Datasets

In this section we will describe the datasets used in our experiments. For this work, we developed a custom suite package to create a diversified set of Ising grid images. These images were generated using the single-flip Metropolis-Hastings algorithm with random temperature variations (between 0K-4K) and three different boundary conditions: **(a)** Periodic boundary conditions (ferromagnetic and anti-ferromagnetic systems): Converts the plane square lattice system into a torus lattice, **(b)** Anti-periodic boundary conditions (both directions): The sign of the coupling constant is reversed at the top/bottom and left/right boundaries of the lattice grid, and **(c)** Skewed ± boundary conditions: Imposes periodic boundary conditions on a lattice except for at the top and bottom rows.

The experiments for the Ising model are carried out on a 100x100 two dimensional lattice. Each lattice point allows for either spin up or spin down, hence there exist $2^{100\text{x}100}$ potential configurations.

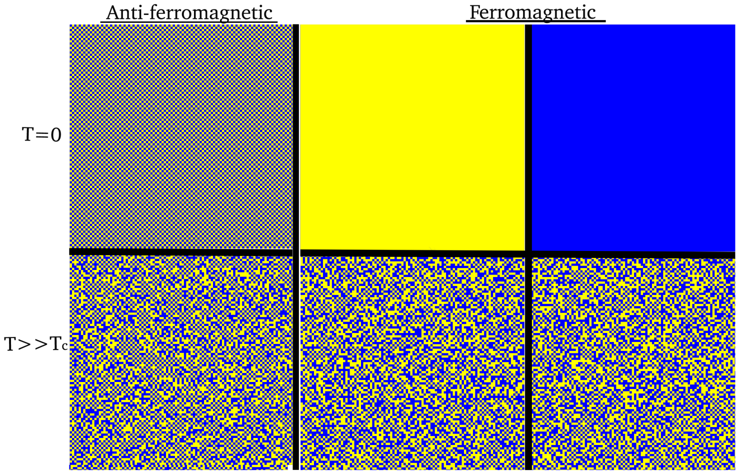

Figure 1: Images of ferromagnetic and anti-ferromagnetic at T = 0 (distinct ground state energy) and T = 4 (disordered system)

We ensure equilibrium, or thermalization is reached between temperature steps by only collecting data after 500 Monte Carlo sweeps. So we wait 750 (cushion for equilibrium) Monte Carlo sweeps prior to the subsequent calculation [22]. A strong proxy for determining whether or not equilibrium has been reached at a new temperature is elucidated by plotting the average magnetization per spin after each implementation of the Metropolis algorithm against the number of iterations.

We generated 1300 images for each boundary condition with step size increments of 0.01K. These images were classified into 4 bins of dissimilar sizes. It is important to point out that for temperatures far above the critical temperature ($T >> 2.27$K), the models tend toward purely noisy systems. The unequal bin sizes were selected to show the effectiveness of the classifier in the small intervals around the critical temperature, and to demonstrate the reliability of predictions in qualitatively differentiable subregions: 0K-1.055K which we call the **far sub-critical region** (FSbCR); 1.055K-2.119K deemed the **sub-critical region** (SbCR); 2.119K-2.320K as the **critical region** (CR); and 2.320K-4.0K as the **super-critical region** (SpCR). Bins were selected to show the ability/sensitivity of the model to identify images in a small neighborhood around the critical temperature (2.27K). This sensitivity is demonstrated in the fact that the size of critical region bin is deliberately selected to be at most 30% the size of the other bins. We are interested in developing a ML model that can accurately predict the label for the image since we believe that such a system can potentially identify characteristic phenomena associated with known critical regions. Interpreting such a model can shed light on what features are associated with multiple critical regions, thus allowing us to look for similar ones in other systems. An example of our dataset is in Figure 1 and 2, whereas the data distribution for various bins are in Table 1. For more figures, please see Appendix.

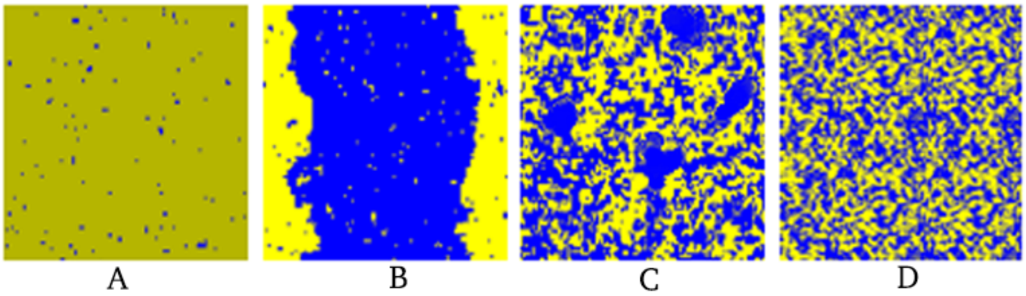

Figure 2: Various Normalized Images Representing Micro-states. The blue area correspond to "spin up" region and the yellow area correspond to "spin down" region. Fig A, B, C, D belongs to FSbCR, SbCR, Cr, SpCR classes respectively.

4 Methods

In this section we will describe our method. We start by giving a short introduction to the Vision Transformer and describe how we train the ViT model for our downstream classification task.

4.1 Vision Transformers

Transformers have become the de-facto sequence-to-sequence architecture in NLP after their introduction in the landmark paper [23]. However, attention-based mechanisms in CV have only been used in conjunction with CNNs. CNNs have been used extensively in CV, as they enjoy inductive biases like translation invariance and a locally restricted receptive field. However, Dosovitskiy et al. [24] successfully introduced Vision Transformer, a transformer-based architecture, to CV which outperforms state-of-the-art CNNs on various downstream CV tasks. Transformers lack the inductive biases of CNNs, and by design they cannot process grid-structured data. The authors cleverly convert a spatial, non-sequential signal by splitting an image into a sequence of flattened, non-overlapping patches. The success of ViT in CV has led to an explosion of research in creating parameter efficient transformers [25], and faster training of self-supervised transformers [26, 27].

In our work, we used the ViT-Base model [24] from the Hugging Face library [28]. We used PyTorch and the Hugging Face Library to train our models on a NVIDIA V100 16GB GPU.

4.2 Finetuning Vision Transformers

We started with the checkpoint of ViT which was pre-trained on ImageNet-21k (a collection of 14 million images and 21k classes) and further fine-tuned on ImageNet (a collection of 1.3 million images and 1,000 classes). We then fine-tuned this ViT model on our training set for 10 epochs with cross-entropy loss. We used AdamW optimizer [29] with learning rate 2e-5 and a weight decay of 5e-6, and early stopping on the validation loss to prevent overfitting. To benchmark our performance, we also fine-tuned a ResNet-18 and ResNet-50 on our data. For a fair comparison, both the ResNets were pretrained on ImageNet-21k and then fine-tuned on ImageNet.

5 Results

In this section, we will describe the performance of our model. Table 2 shows that the ViT outperforms both the ResNets when fine-tuned on such a small dataset. This is an active work in progress as we benchmark ViT on other Ising model simulations like the Blume-Capel model and the q-Potts model.

System/Boundary Conditions	ViT	ResNet-18	ResNet-50
Periodic (Ferromagnetic)	**0.934 ± .008**	0.865 ± .034	0.907 ± .021
Periodic (Anti-ferromagnetic)	**0.935 ± .012**	0.906 ± .016	0.899 ± .026
Skewed (Ferromagnetic)	**0.931 ± .021**	.886 ± .019	0.917 ± .009
Anti-Periodic (Ferromagnetic)	**0.921 ± .013**	0.91 ± .021	0.917 ± .008

Table 2: Table showing the Macro F1 scores (average and standard deviation over 5 trials) of various models on our test sets

6 Conclusion

We developed a custom suite to generate images from various Ising model simulations at different temperatures and with different boundary conditions. We then used these images to fine-tune a pre-trained ViT and demonstrated that ViT can outperform the current state-of-the-art CNNs. This is still a work in progress as we extend our work to other simulations and applications; to systems undergoing topological phase transition, e.g. XY model, systems widely used to model behaviors of complex systems, e.g. q-Potts model, and the transverse field quantum Ising model (continuous imaginary-time). One of our main motivations in using ViT is to use the attention maps for interpretability. Our goal is to use the attention maps to understand Ising model phase transitions from a visual pattern perspective by looking at Ising configurations, with the different boundary conditions generating qualitatively different looking configurational patterns. We believe this will raise interesting research questions as we try to relate conventional physical approaches usually applied to the Ising problem with the new ideas in computer vision.

7 Broader Impact

The past year has produced a unprecedentedly rapid and cross-disciplinary increase in the study and applications of CV systems across almost every corner of life and industry. Much of the interest is due to the extensive and seemingly endless applicability of transformers and other state of the art deep learning frameworks. In particular, our work shows that state-of-the-art research in machine learning can be used in understanding fundamental sciences and potentially help in discovering new natural phenomena. However we must be mindful of harmful effects of using these computer vision technologies in society. There are well known examples where these computer vision technologies has mischaracterized people of color and sometimes it has led to severe consequences like harassment towards Black and Brown people. Situations like this can be mitigated with careful focus at each step of the development process. Moreover more research needs to be done in interpreting these models and at the end of the day we should acknowledge that our methods in understanding the predictions of these black box models are incomplete. Thus we must be mindful of how we use technology when we wish to introduce new, powerful, and exciting technologies into society.

Acknowledgments and Disclosure of Funding

We would like to thank Christina Tang (contact: christina.f.tang@gmail.com) for technical assistance and for refactoring our Ising model simulations suite code into a fully customizable experimentation suite. We would like to thank Tamiko Jenkins (contact: tommymjenkins@gmail.com) for her tireless efforts and input while proof–reading and editing this manuscript. We would also like to thank John Weeks and Yigit Subasi for helpful comments and suggestions on an earlier version of the manuscript.

References

[1] Ernst Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, Feb 1925.

[2] Lars Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65:117–149, Feb 1944.

[3] Leo Radzihovsky. Introduction to quenched disorder. *Boulder Summer School for Condensed Matter and Materials Physics*, Soft Matter In and Out of Equilibrium:1–32, 2015.

[4] W.A. Little. The existence of persistent states in the brain. *Mathematical Biosciences*, 19(1):101–120, 1974.

[5] J. J. Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences of the United States of America*, 79(8):2554–2558, Apr 1982. PMC346238[pmcid].

[6] E Gardner. The space of interactions in neural network models. *Journal of Physics A: Mathematical and General*, 21(1):257–270, jan 1988.

[7] Kenta Shiina, Hiroyuki Mori, Yutaka Okabe, and Hwee Kuan Lee. Machine-learning studies on spin models. *Scientific Reports*, 10(1), Feb 2020.

[8] Antoine Dedieu, Miguel Lázaro-Gredilla, and Dileep George. Sample-efficient l0-l2 constrained structure learning of sparse ising models, 2021.

[9] Ahmadreza Azizi and Michel Pleimling. A cautionary tale for machine learning generated configurations in presence of a conserved quantity. *SCIENTIFIC REPORTS*, 11(1), MAR 18 2021.

[10] Alan Morningstar and Roger G. Melko. Deep learning the ising model near criticality. *Journal of Machine Learning Research*, 18(163):1–17, 2018.

[11] Juan Carrasquilla and Roger G. Melko. Machine learning phases of matter. *Nature Physics*, 13(5):431–434, May 2017.

[12] Kyle Sprague, Juan Carrasquilla, Stephen Whitelam, and Isaac Tamblyn. Watch and learn—a generalized approach for transferrable learning in deep neural networks via physical principles. *Machine Learning: Science and Technology*, 2(2):02LT02, Feb 2021.

[13] Kimihiko Fukushima and Kazumitsu Sakai. Can a cnn trained on the ising model detect the phase transition of the q-state potts model? *Progress of Theoretical and Experimental Physics*, 2021(6), May 2021.

[14] Stavros Efthymiou, Matthew J. S. Beach, and Roger G. Melko. Super-resolving the ising model with convolutional neural networks. *Physical Review B*, 99(7), Feb 2019.

[15] Akinori Tanaka and Akio Tomiya. Detection of phase transition via convolutional neural networks. *Journal of the Physical Society of Japan*, 86(6):063001, Jun 2017.

[16] Constantia Alexandrou, Andreas Athenodorou, Charalambos Chrysostomou, and Srijit Paul. The critical temperature of the 2d-ising model through deep learning autoencoders. *The European Physical Journal B*, 93(12):226, Dec 2020.

[17] S. Acevedo, M. Arlego, and C. A. Lamas. Phase diagram study of a two-dimensional frustrated antiferromagnet via unsupervised machine learning. *Phys. Rev. B*, 103:134422, Apr 2021.

[18] Dimitrios Bachtis, Gert Aarts, and Biagio Lucini. Mapping distinct phase transitions to a neural network. *Phys. Rev. E*, 102:053306, Nov 2020.

[19] Zhaocheng Liu, Sean P. Rodrigues, and Wenshan Cai. Simulating the ising model with a deep convolutional generative adversarial network, 2017.

[20] Nicholas Walker, Ka-Ming Tam, and Mark Jarrell. Deep learning on the 2-dimensional ising model to extract the crossover region with a variational autoencoder. *Scientific Reports*, 10(1):13047, Aug 2020.

[21] Francesco D'Angelo and Lucas Böttcher. Learning the ising model with generative neural networks. *Phys. Rev. Research*, 2:023266, Jun 2020.

[22] David P. Landau and Kurt Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 4 edition, 2014.

[23] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

[24] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16× 16 words: Transformers for image recognition at scale, 2020.

[25] Haiping Wu, Bin Xiao, Noel Codella, Mengchen Liu, Xiyang Dai, Lu Yuan, and Lei Zhang. Cvt: Introducing convolutions to vision transformers, 2021.

[26] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers, 2021.

[27] Andreas Steiner, Alexander Kolesnikov, Xiaohua Zhai, Ross Wightman, Jakob Uszkoreit, and Luca Beye. How to train your vit? data, augmentation, and regularization in vision transformers, 2021.

[28] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. Huggingface's transformers: State-of-the-art natural language processing, 2020.

[29] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization, 2019.

A Appendix

A.1 Monte Carlo Sampling Algorithm

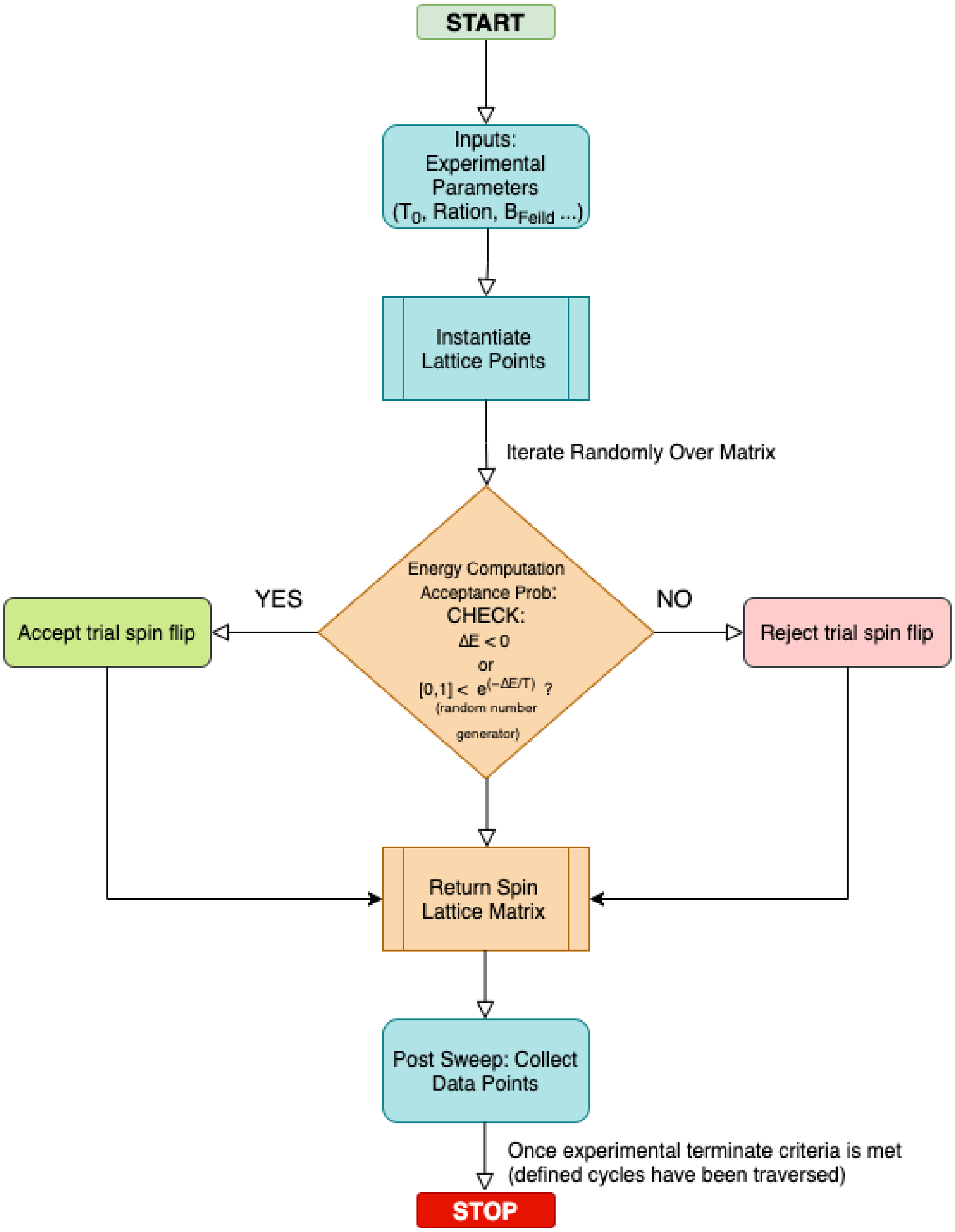
START
Inputs: Experimental Parameters (T_0, Ration, B_{Felld} ...)
Instantiate Lattice Points
Iterate Randomly Over Matrix
Energy Computation
Acceptance Prob:
CHECK:
$\Delta E < 0$
or
$[0,1] < e^{(-\Delta E/T)}$?
(random number generator)
YES
NO
Accept trial spin flip
Reject trial spin flip
Return Spin Lattice Matrix
Post Sweep: Collect Data Points
Once experimental terminate criteria is met
(defined cycles have been traversed)
STOP

Figure 3: Flow-chart depicting a single sweep though the Metropolis algorithm

A.2 Normalized Micro-state Image Panels

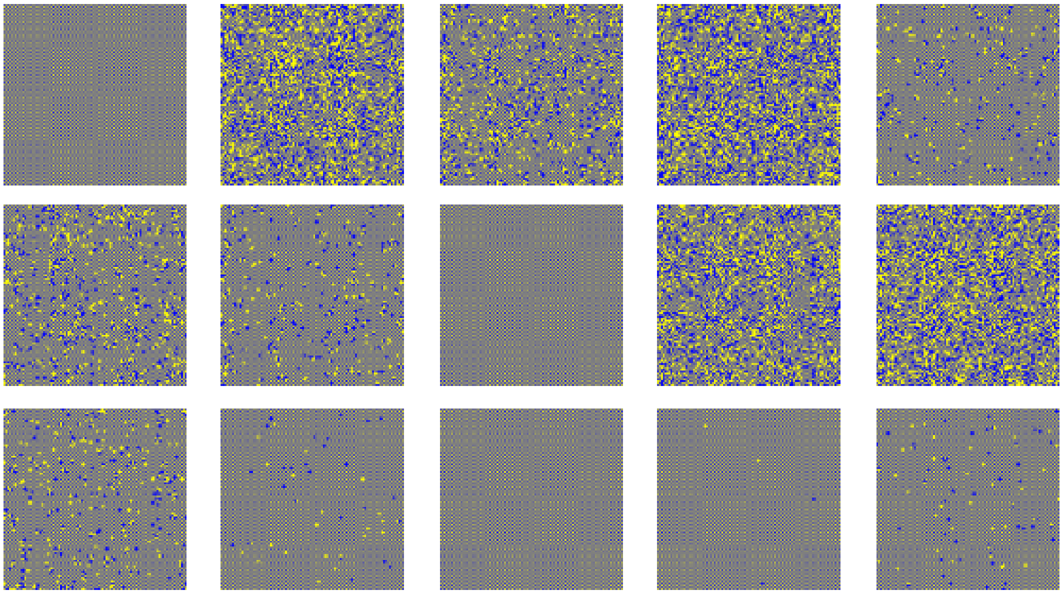

Figure 4: Anti-Ferromagnetic Ising Model With Periodic Boundary Conditions

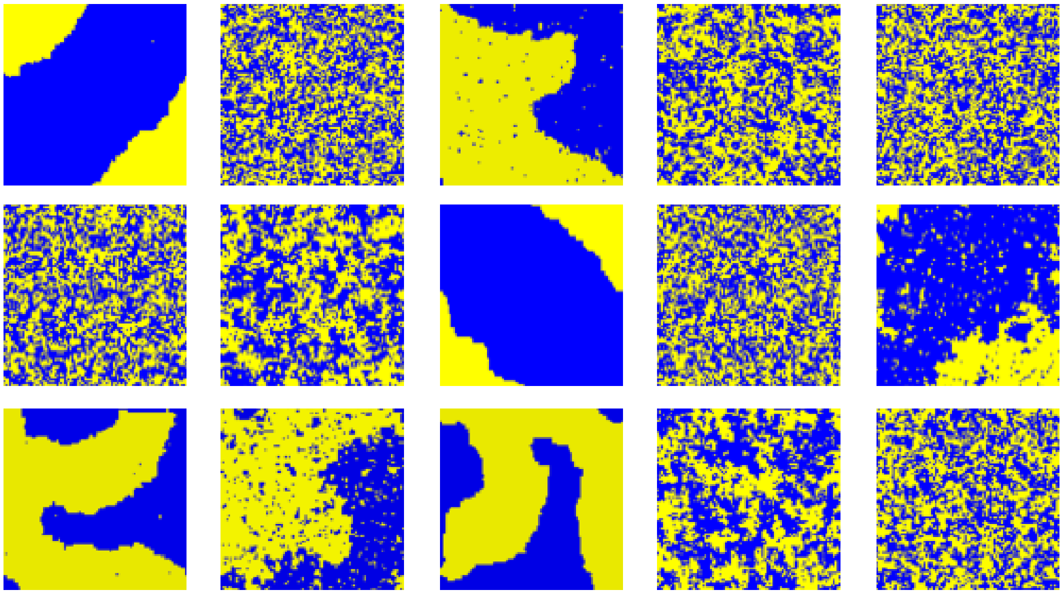

Figure 5: Ising Model With Anti-Periodic Boundary Conditions

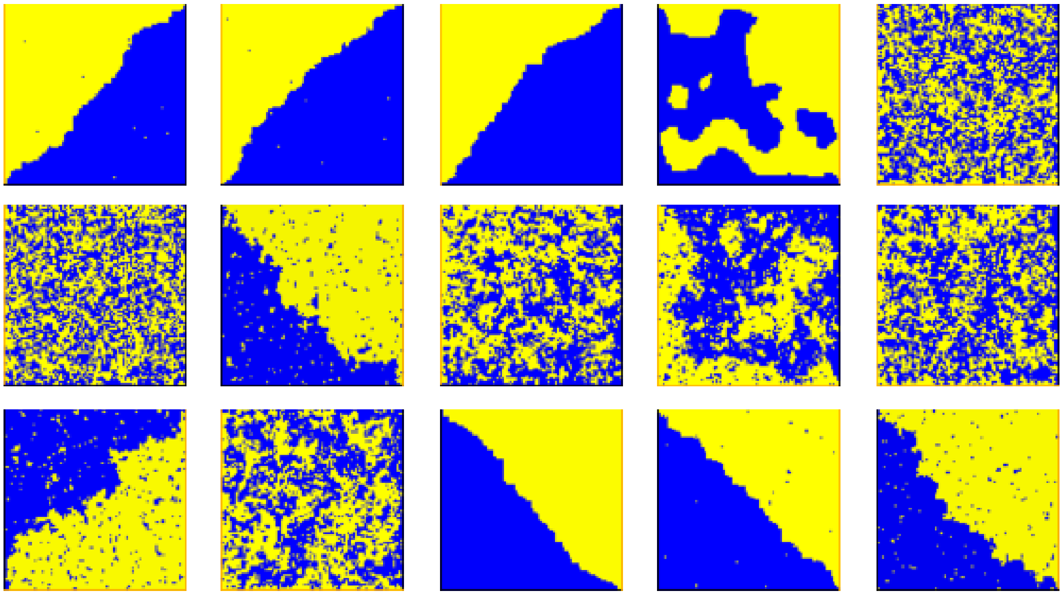

Figure 6: Ising Model With Skewed ± Boundary Conditions

Checklist

The checklist follows the references. Please read the checklist guidelines carefully for information on how to answer these questions. For each question, change the default **[TODO]** to [Yes] , [No] , or [N/A] . You are strongly encouraged to include a **justification to your answer**, either by referencing the appropriate section of your paper or providing a brief inline description. For example:

- Did you include the license to the code and datasets? [Yes] The code will be made publicly available on publication

Please do not modify the questions and only use the provided macros for your answers. Note that the Checklist section does not count towards the page limit. In your paper, please delete this instructions block and only keep the Checklist section heading above along with the questions/answers below.

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? [Yes] We describe our contributions in the Introduction 1 and in the Conclusion 6.
   (b) Did you describe the limitations of your work? [Yes] Our methods do not generalize to the 3D Ising model. More crucially we understand that attention maps may not be the best tool to interpret our models.
   (c) Did you discuss any potential negative societal impacts of your work? [N/A]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [N/A]
   (b) Did you include complete proofs of all theoretical results? [N/A]
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [No] We are in process of cleaning up the code and creating a python package. We will publish the package as a part of a long paper currently under review at a journal.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes] In the section Finetuning Vision Transformers 4.2.
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [Yes] Yes in the section Vision Transformers 4.1.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [N/A]
   (b) Did you mention the license of the assets? [N/A]
   (c) Did you include any new assets either in the supplemental material or as a URL? [N/A]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]