Learning Renormalization-Group-Like Latent Variables with a Hierarchical GNN-VAE for the 2D Ising Model (Extended Abstract)

Tingyu Meng
University of Wisconsin–Madison

Abstract

We study whether a generative, self-supervised neural network can learn *renormalization-group (RG)-like* variables from raw Monte Carlo configurations of the two-dimensional Ising model. Our model is a hierarchical graph neural network variational autoencoder (GNN-VAE) equipped with an *RG consistency* loss that encourages the latent representation to remain stable under repeated coarse-graining. We find that the learned latent space (i) correlates with magnetization and energy density, (ii) exhibits two basins consistent with ordered/disordered fixed points, and (iii) supports an empirical linear "RG map" whose dominant eigen-direction aligns with the leading principal component (PC1) of the latent space. These observations suggest that the network learns an interpretable, scale-robust coordinate beyond mere phase separation.

1 Motivation

The RG provides a unifying description of critical phenomena by mapping a microscopic model to an effective theory at longer length scales [1, 2]. A key question in modern ML-for-physics is whether neural networks can *discover* such coarse variables directly from data rather than being used solely as phase classifiers [3, 4]. We focus on a generative setting (VAE [5]) and introduce an explicit inductive bias—*latent scale consistency*—to encourage RG-like structure.

2 Model and Training

We generate 2D Ising configurations on an $L \times L$ square lattice ($L = 16$) with periodic boundary conditions using Metropolis updates. Each configuration is encoded as a lattice graph with nearest-neighbor edges. A tokenizer augments each spin $s_i \in \{-1, +1\}$ with a local interaction feature $h_i = \sum_{j \in \text{n.n.}(i)} s_i s_j$ before message passing.

Our encoder is hierarchical: repeated GCN+pooling blocks perform coarse-graining $16 \to 8 \to 4$ and output latent parameters $(\mu^{(\ell)}, \log \sigma^{2(\ell)})$ at each level. A graph-level latent is sampled via the reparameterization trick

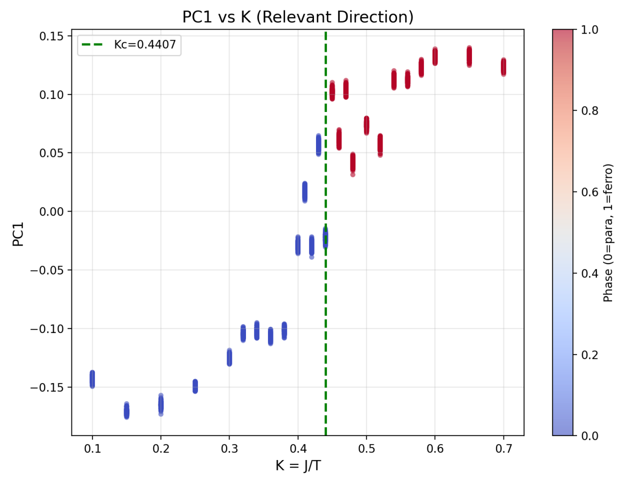

Figure 1: First principal component (PC1) of the latent mean μ versus coupling $K = J/T$. PC1 changes most rapidly near $K_c \simeq 0.4407$, consistent with a relevant thermal direction organizing the two phases.

$z^{(\ell)} = \mu^{(\ell)} + \sigma^{(\ell)}\epsilon$, and an MLP decoder reconstructs the spin configuration from the final latent.

The objective combines reconstruction, KL regularization, and an RG consistency term:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + \beta\, D_{\text{KL}}(q(z|x) \,\|\, \mathcal{N}(0, I)) + \lambda_{\text{RG}} \mathcal{L}_{\text{RG}}, \quad \mathcal{L}_{\text{RG}} = \frac{1}{L_s - 1} \sum_{\ell} \| \tag{1}$$

where $r^{(\ell)}$ is a latent representation at level ℓ (in practice we use $\mu^{(\ell)}$ for stability) and L_s is the number of scales.

3 Key Results

Latent interpretability. Performing PCA on the learned latent means, PC1 varies systematically with the coupling K and correlates with standard observables (magnetization and energy density), indicating that the dominant variance direction is physically meaningful (Fig. 1).

RG-like flow and fixed points. Tracking the same configuration under repeated coarse-graining in latent space reveals trajectories that flow toward two distinct basins, consistent with ordered/disordered fixed points.

Empirical RG map. We fit a linear map between latents at adjacent scales, $\mu_{\text{coarse}} \approx W\mu_{\text{fine}} + b$, and analyze the spectrum of W. The dominant eigen-direction aligns closely with PC1 (cosine similarity ≈ 1 in our runs), supporting the interpretation of PC1 as the learned "relevant" direction. Excluding samples near the critical region stabilizes the fitted spectrum, highlighting the sensitivity of linearization near criticality.

4 Discussion and Deliverables

Our results provide multiple, complementary signatures of RG-like organization in a learned latent space: scale-consistent representations, physically interpretable latent axes, and agreement between PCA and a fitted linear RG map. Remaining work toward a publication-quality result includes controlled ablations (e.g., $\lambda_{\text{RG}} = 0$) and likelihood choices appropriate for binary spins (Bernoulli/BCE instead of MSE).

Code and repository link. Please upload your GitHub URL here: <YOUR_GITHUB_REPO_URL>. The analysis figures are generated by `analyze_all.py` and saved in `analysis_plots/`. The implementation uses PyTorch and PyTorch Geometric [6, 7].

References

References

[1] Kenneth G. Wilson. Renormalization group and critical phenomena. I. renormalization group and the kadanoff scaling picture. *Phys. Rev. B*, 4:3174–3183, 1971.

[2] Leo P. Kadanoff. Scaling laws for ising models near t_c. *Physics*, 2:263–272, 1966.

[3] Lei Wang. Discovering phase transitions with unsupervised learning. *Phys. Rev. B*, 94:195105, 2016.

[4] Sebastian J. Wetzel. Unsupervised learning of phase transitions: From principal component analysis to variational autoencoders. *Phys. Rev. E*, 96:022140, 2017.

[5] Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2014.

[6] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in Neural Information Processing Systems*, volume 32, 2019.

[7] Matthias Fey and Jan Eric Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.

Appendix (optional): additional figures

Note: Appendix pages do not count toward the 2-page limit.

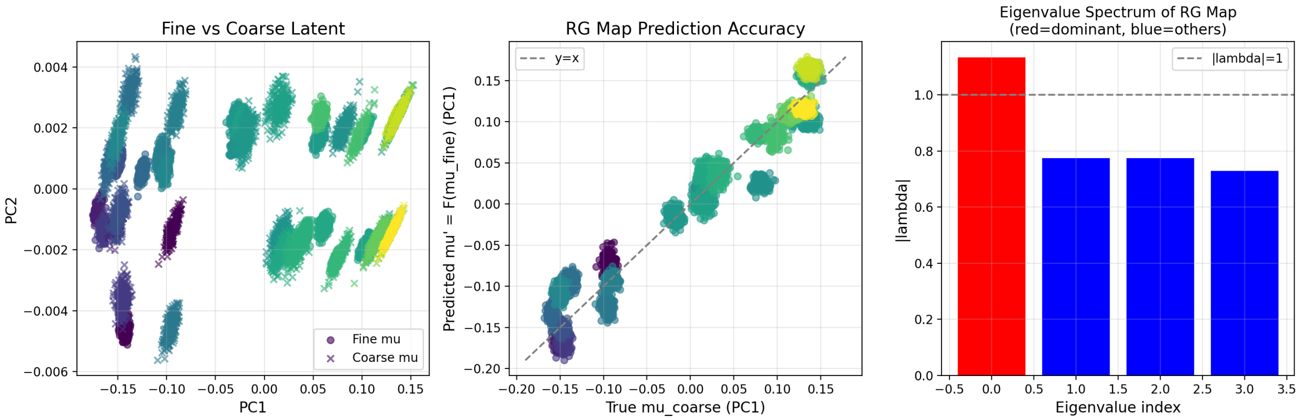

Figure 2: RG map analysis: fine vs coarse latent overlap, linear map accuracy, and eigenvalue spectrum (see main text).

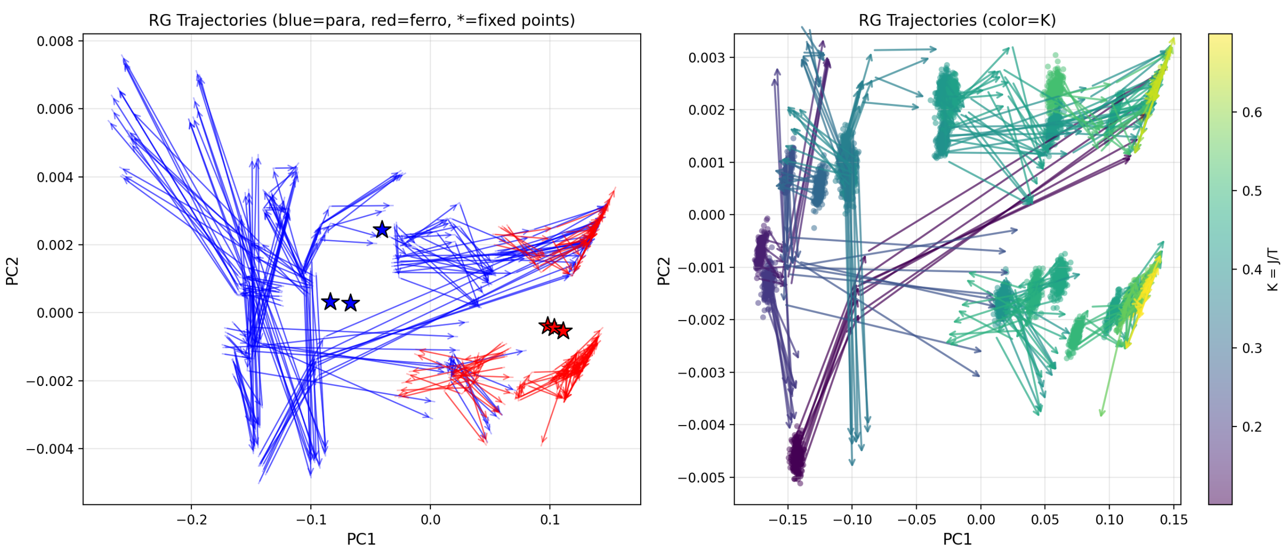

Figure 3: Latent RG trajectories under repeated coarse-graining (fine→coarse).