Renormalization Group and Generative Modeling

Yi-Zhuang You (尤亦庄)
University of California, San Diego

Lattice Field Theory and Machine Learning

Taiwan, Dec. 2024

Machine Learning and Physics

- Bidirectional objective:
  - Can ML techniques help us discover physical laws?
  - Can physics ideas help us develop ML algorithms?

- RG-Flow: a hierarchical *flow-based generative model* motivated by the *renormalization group* in physics.
  - Application 1: simulating critical systems

    [1] H Hu, SH Li, L Wang, YZ You. arXiv: **1903.00804**

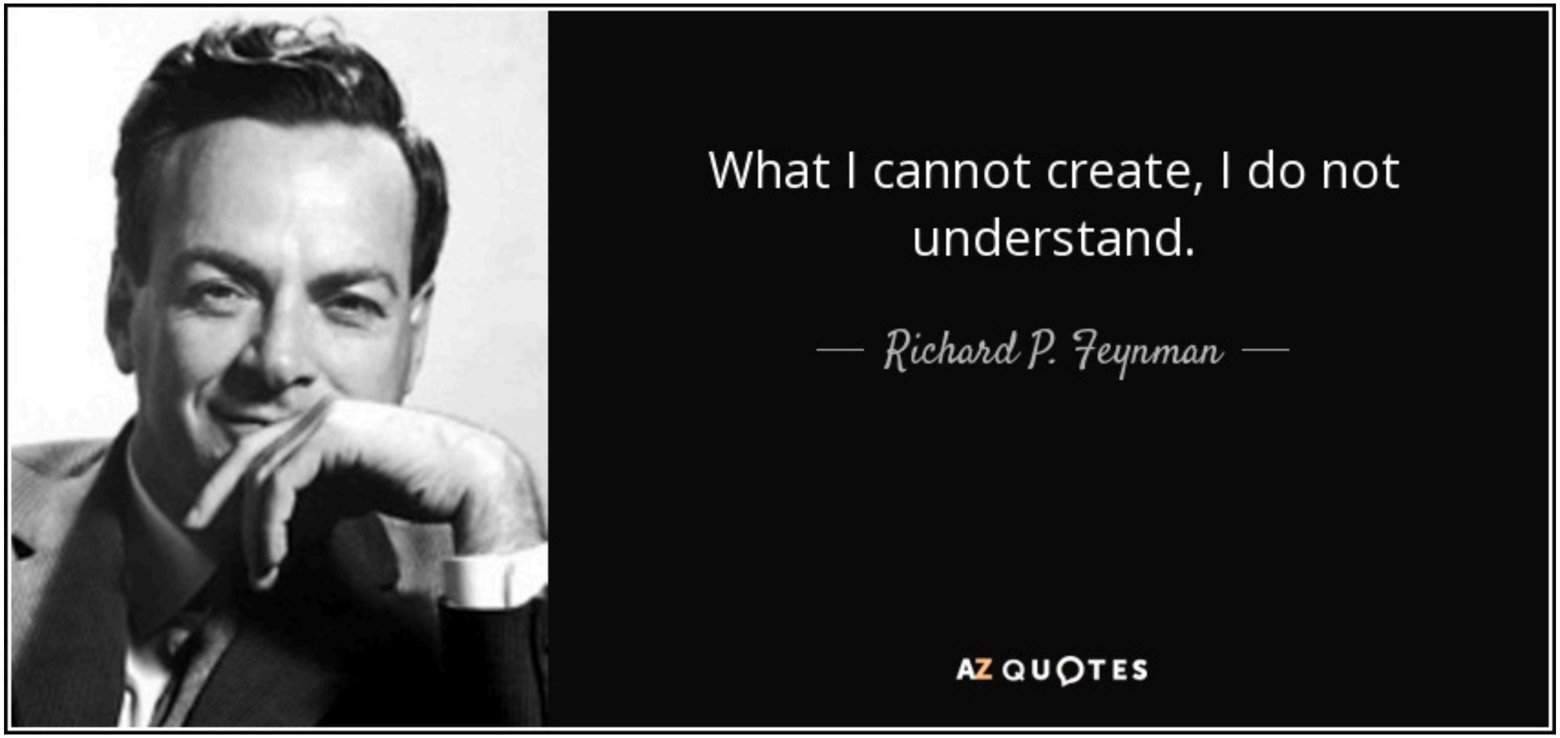
  - Application 2: image processing.

    [2] H Hu, D Wu, YZ You, B Olshausen, Y Chen. arXiv: **2010.00029**

    [3] A Sheshmani, YZ You, W Fu, A Azizi. arXiv: **2203.07975**

Generative Modeling

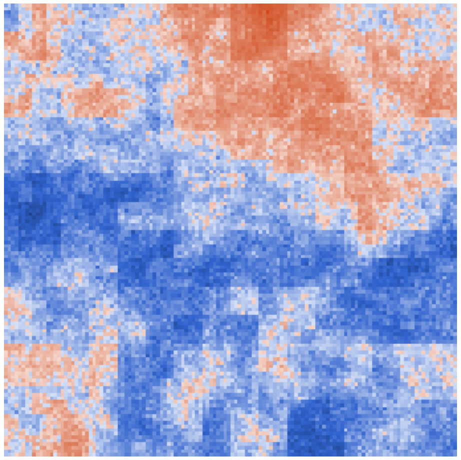
- Generative modeling (unsupervised learning) is an important topic in machine learning.
- It aims to *model* the **probability distribution** of samples in the dataset and *create* new samples based on the learned distribution.



Generative Model and Quantum Field Thoery

- Quantum field theory = generative model of quantum fields
  - Sample: field configuration

$$\boldsymbol{x} =$$



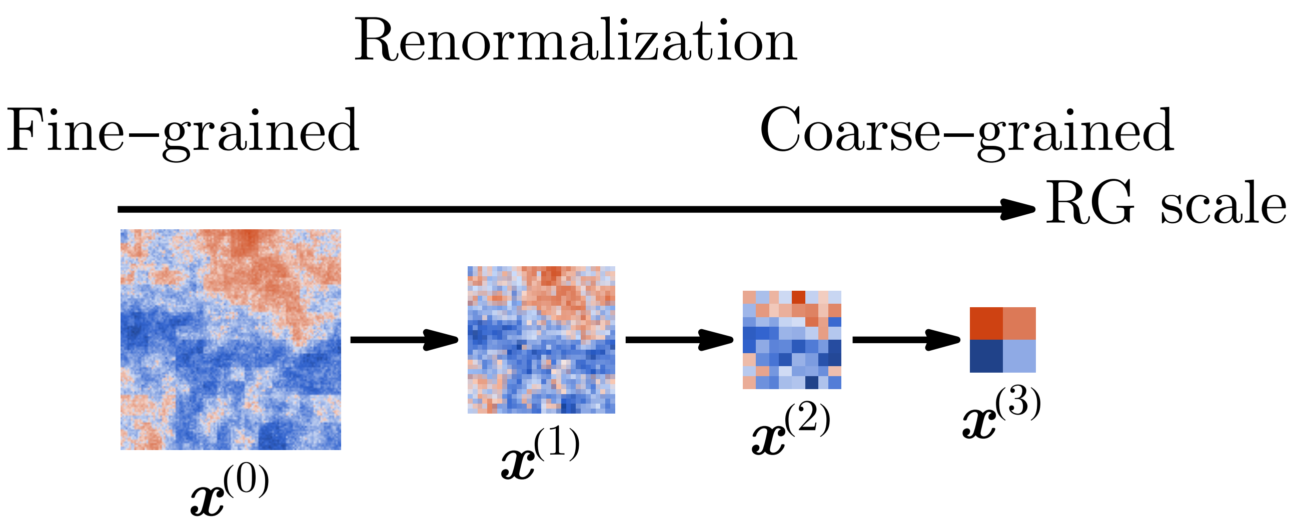
  - Negative log likelihood: field action (energy function)

$$p(\boldsymbol{x}) \propto \mathrm{e}^{-S(\boldsymbol{x})}$$

- The renormalization group is an important approach to analyze quantum field theory, which systematically extracts the effective action at different scales.

Renormalization Group and Deep Learning

- The similarity between the renormalization group (RG) and deep learning has long been noticed.
- RG transformation (coarse-graining rule)

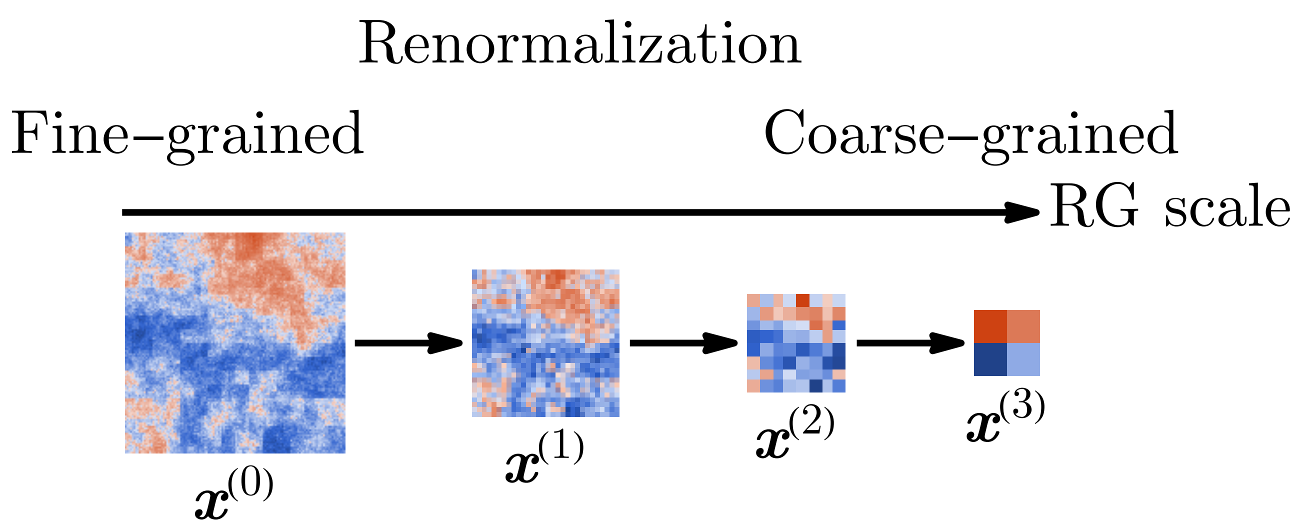


$$p(\boldsymbol{x}^{(l+1)}|\boldsymbol{x}^{(l)})$$

C Beny, arXiv: 1301.3124. P Mehta, DJ Schwab, arXiv: 1410.3831. HW Lin, M Tegmark, D Rolnick, arXiv: 1608.08225. EdM Koch, RdM Koch, L Cheng, arXiv: 1906.05212. JH Chung, YJ Kao, arXiv: 2010.05703 …

Renormalization Group and Deep Learning

- The RG transformation induces a flow of the underlying probability model (or the field action)



Fine-grained: $$p(\boldsymbol{x}^{(l)}) \propto \mathrm{e}^{-S(\boldsymbol{x}^{(l)})}$$

Coarse-grained: $$p(\boldsymbol{x}^{(l+1)}) = \sum_{\boldsymbol{x}^{(l)}} p(\boldsymbol{x}^{(l+1)}|\boldsymbol{x}^{(l)})p(\boldsymbol{x}^{(l)})$$

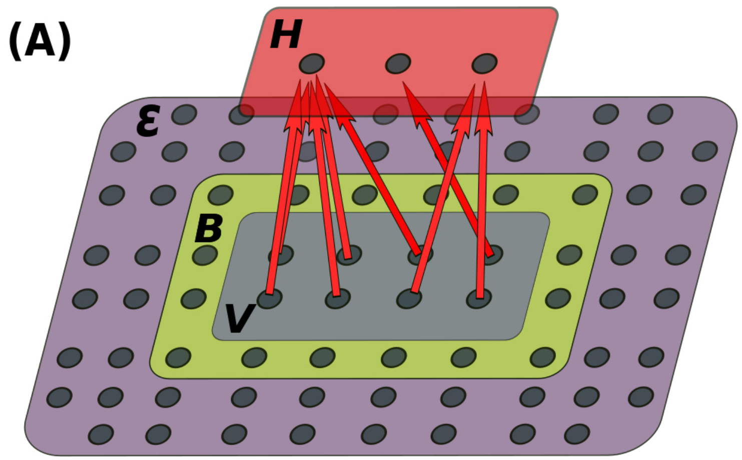
Or $$\mathrm{e}^{-S'(\boldsymbol{x}^{(l+1)})} \propto \sum_{\boldsymbol{x}^{(l)}} p(\boldsymbol{x}^{(l+1)}|\boldsymbol{x}^{(l)})\mathrm{e}^{-S(\boldsymbol{x}^{(l)})}$$

What is the Designing Principle of RG?

- The goal of the renormalization group (RG) is to extract relevant features of field configurations.
- But what should be the relevant feature? The answer can be model/dataset dependent …
  - Real-space RG of Ising models
    - Ferromagnetic coupling: uniform spin component
    - Anti-ferromagnetic coupling: staggered spin component
    - Random coupling: …?
  - Momentum-space RG of field theory
    - Low-energy freedoms are relevant (but what is the notion of “energy” in general?)
- Is there an information-theoretic principle to guide the design of the optimal RG transformation?

What is the Designing Principle of RG?

- Maximal real-space mutual information (maxRMI) principle
  - Relevant features retain maximal mutual information with neighboring environments.



$$\max I(H, \mathcal{E})$$

Coarse-grained freedom

Environment

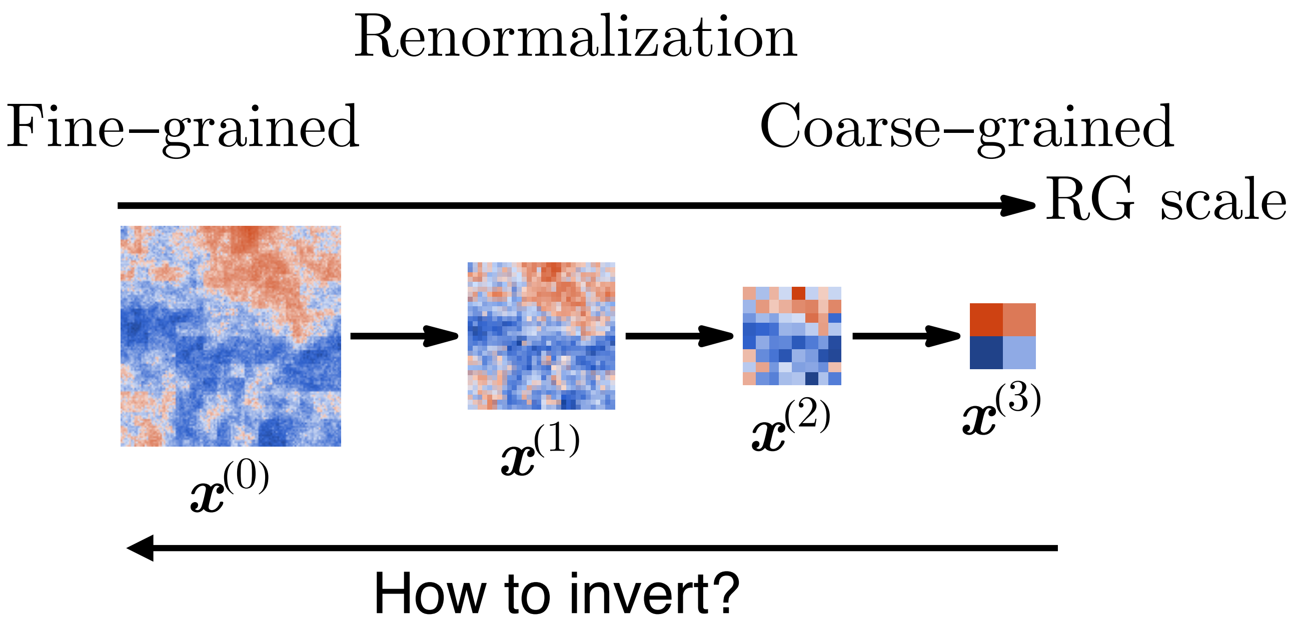
M Koch-Janusz, Z Ringel, arXiv: 1704.06279
A Gordon, A Banerjee, M Koch-Jansz, Z Ringel, arXiv: 2012.01447

  - The RG transform $p(H|V)$ should optimize this objective.
- We proposed another equivalent principle based on invertible RG and holographic mapping.

H Hu, S-H Li, L Wang, Y-Z You. arXiv: 1903.00804

Invertible RG and Holographic Mapping

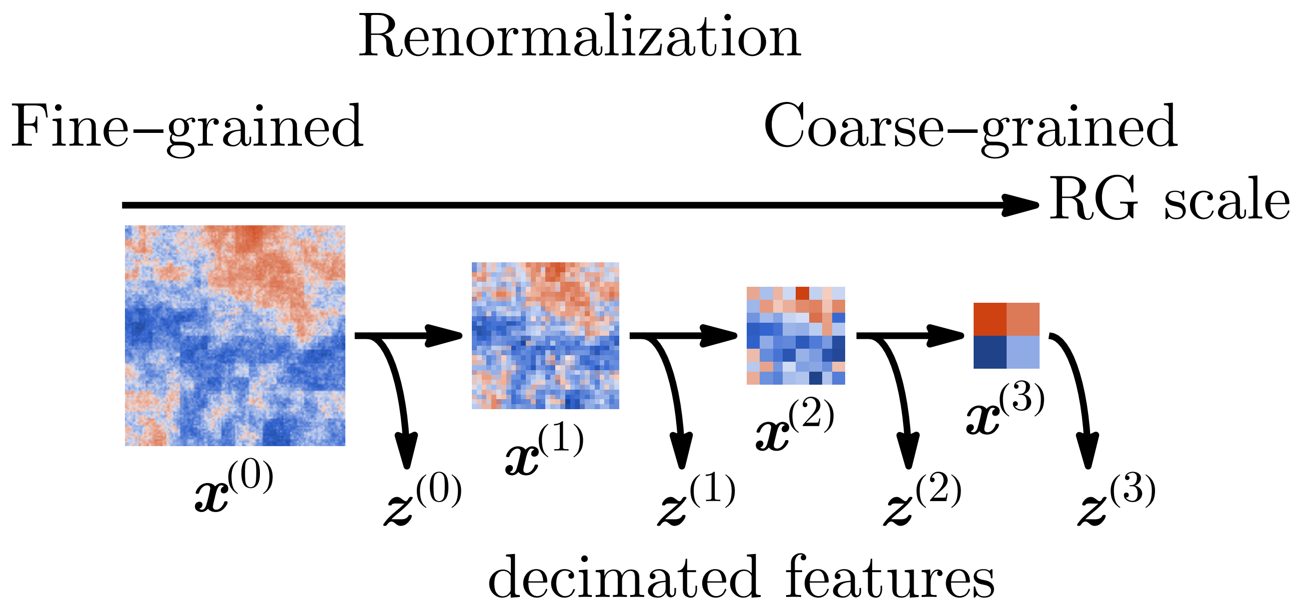
- The renormalization “group” is *not* a group!
  The conventional formulation of RG is information lossy and is therefore irreversible.



$$p(\boldsymbol{x}^{(l+1)}|\boldsymbol{x}^{(l)}) \to p(\boldsymbol{x}^{(l)}|\boldsymbol{x}^{(l+1)})?$$

Invertible RG and Holographic Mapping

- What have been discarded are the irrelevant features at each RG step.

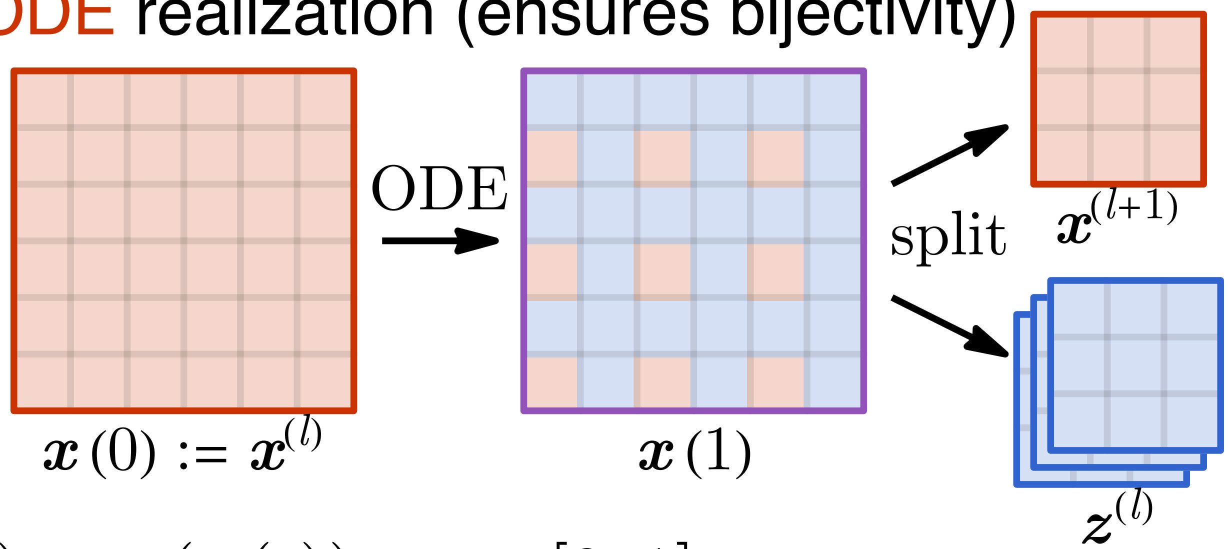


$$\boldsymbol{x}^{(l+1)}, \boldsymbol{z}^{(l)} = R_l(\boldsymbol{x}^{(l)})$$ (Bijective local map)

- $\boldsymbol{x}^{(l+1)}$: Coarse-grained features
- $\boldsymbol{z}^{(l)}$: Irrelevant features to be decimated
- $\boldsymbol{x}^{(l)}$: Fine-grained features

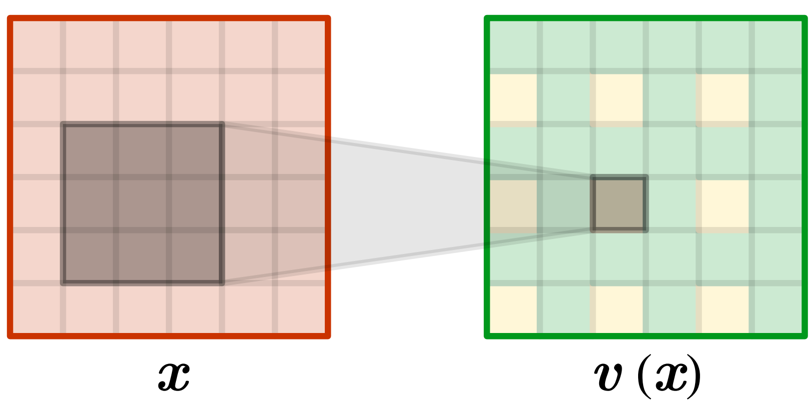
Design of Invertible RG Transform

- Requirement for R: bijectivity + locality (+ equivariance)
- Neural ODE realization (ensures bijectivity)



$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t) = \boldsymbol{v}(\boldsymbol{x}(t)) \quad t \in [0,1]$$

RTQ Chen et.al. arXiv:1806.07366



Velocity model (neural net):

- Impose locality by CNN
- Equivariance can be required

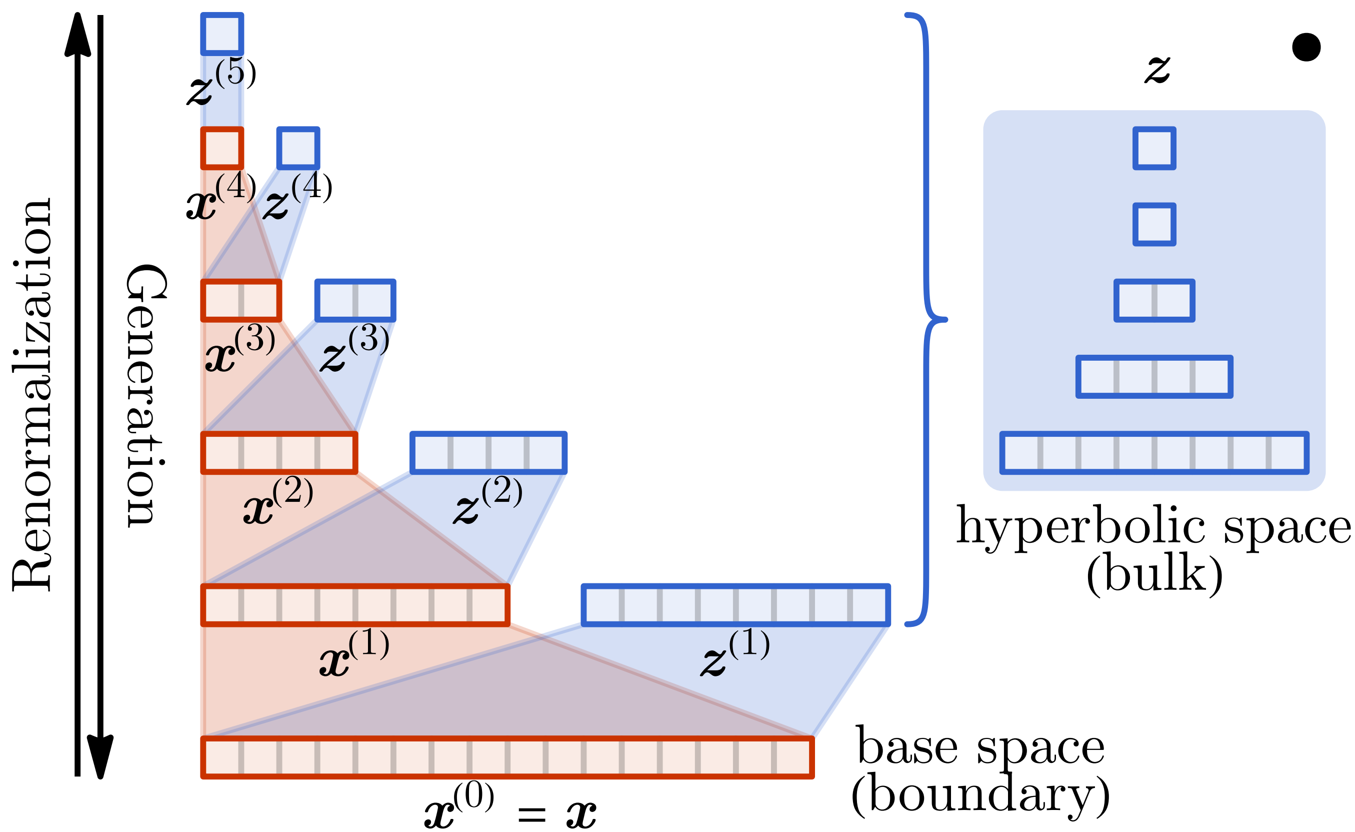
$$G\boldsymbol{v}(\boldsymbol{x}) = \boldsymbol{v}(G\boldsymbol{x})$$

A Sheshmani, YZ You, W Fu, A Azizi. arXiv:2203.07975

Invertible RG and Holographic Mapping

- All the decimated features form a holographic encoding of the original fine-grained configuration.

$$\boldsymbol{z} = (\boldsymbol{z}^{(0)}, \boldsymbol{z}^{(1)}, \cdots) = R(\boldsymbol{x}^{(0)})$$



- Bijective (no information loss)
- Holographic (boundary to bulk)

XL Qi, arXiv: 1309.6282

A Sheshmani, YZ You, W Fu, A Azizi. arXiv:2203.07975

Invertible RG and Holographic Mapping

- Holographic duality (AdS/CFT correpondance): a mapping between a quantum field theory and a gravity theory in one-higher dimension.
- Critical systems (CFT) ↔ hypobolic geometry (AdS)
  - Power-law correlation of boundary variables

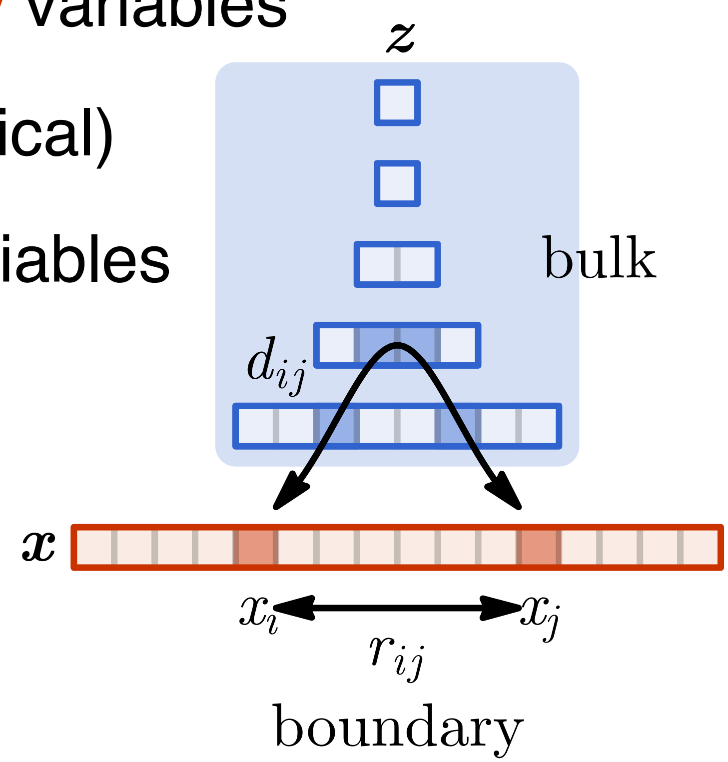
    $$\langle x_i x_j \rangle \sim r_{ij}^{-2\Delta}$$ (Massless, critical)

  - Exponential correlation of bulk variables

    $$\langle z_i z_j \rangle \sim \mathrm{e}^{-2\Delta d_{ij}}$$ (Massive)

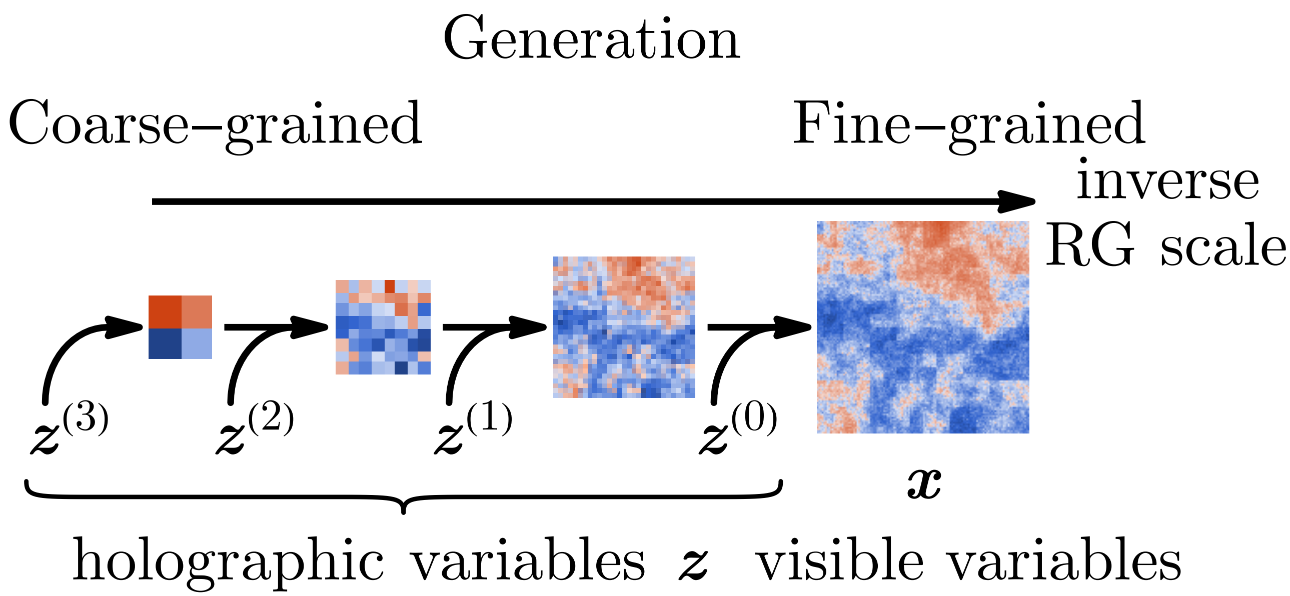
  - Because the boundary and bulk geodesic distances are related by

    $$d_{ij} \sim \ln r_{ij}$$



Invertible RG and Generative Modeling

- Generation is the *inverse* of renormalization!



- One-step inverse map: $x^{(l)} = R_l^{-1}(x^{(l+1)}, z^{(l)})$
- Holographic decoding (bulk to boundary)

$$x := x^{(0)} = R_0^{-1}(R_1^{-1}(R_2^{-1}(\cdots, z^{(2)}), z^{(1)}), z^{(0)}) := R^{-1}(z)$$

Invertible RG and Generative Modeling

- The holographic mapping R is deterministic. How can it be used to model probability distribution?
- Flow-based generative model: deforming a known *prior* distribution to the *target* distribution by *bijective* maps.

	Dual	
$\boldsymbol{x} = R^{-1}(\boldsymbol{z})$	⟷	$\boldsymbol{z} = R(\boldsymbol{x})$
$p_X(\boldsymbol{x}) = p_Z(\boldsymbol{z}) \det\left(\frac{\partial \boldsymbol{z}}{\partial \boldsymbol{x}}\right)$		$p_Z(\boldsymbol{z}) = p_X(\boldsymbol{x}) \det\left(\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{z}}\right)$
Target Prior		Prior Target

- What to learn?
  - The bijective RG transformation R
- How to learn?
  - Given prior match target $\mathcal{L} = D_{\mathrm{KL}}(p_{\mathrm{dat}}(\boldsymbol{x}) \| p_X(\boldsymbol{x}))$
  - Or given target match prior $\mathcal{L} = D_{\mathrm{KL}}(p_{\mathcal{N}}(\boldsymbol{z}) \| p_Z(\boldsymbol{z}))$

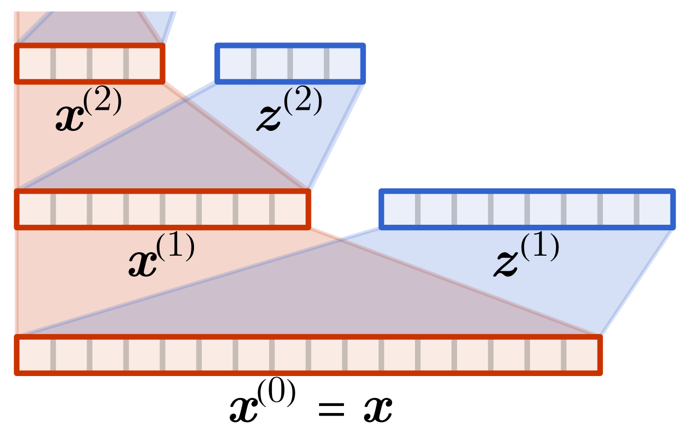
L Dinh, J Sohl-Dickstein, S Bengio, arXiv: 1605.08803.

What is the Designing Principle of RG?

- Maximal real-space mutual information (maxRMI) principle:
  - Relevant features retain maximal mutual information with neighboring environments.
- Minimal bulk mutual information (minBMI) principle:
  - Irrelevant features have minimal mutual information with each other.

$$\min I(z_i^{(l)}, z_j^{(l')})$$

H Hu, S-H Li, L Wang, Y-Z You. arXiv: 1903.00804



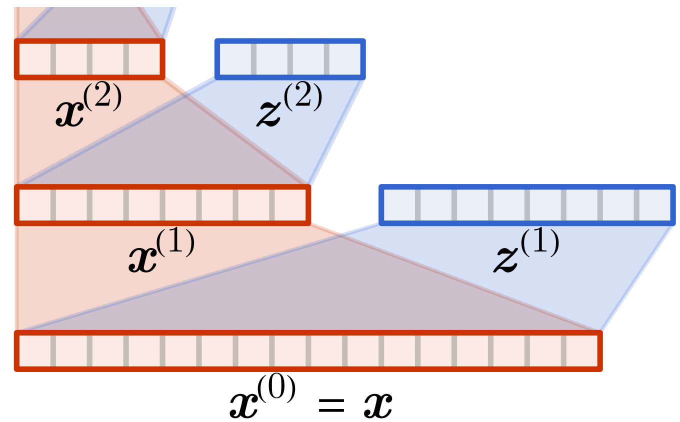
  - Holographic interpretation: the bulk freedom should be massive.
  - RG flow is the optimal transport of a non-trivial field theory (boundary) towards a massive Gaussian theory (bulk).

J Cotler, S Rezchikov, arXiv:2202.11737

What is the Designing Principle of RG?

- Maximal real-space mutual information (maxRMI) principle:
  - Relevant features retain maximal mutual information with neighboring environments.
- Minimal bulk mutual information (minBMI) principle:
  - Irrelevant features have minimal mutual information with each other.

$$\min I(z_i^{(l)}, z_j^{(l')})$$



  - Objective function

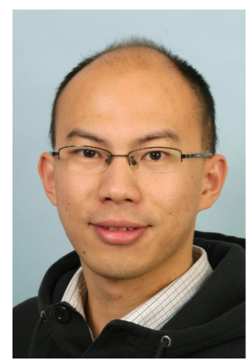
$$\mathcal{L} = D_{\mathrm{KL}}(p_{\mathcal{N}}(\boldsymbol{z}) \| p_Z(\boldsymbol{z}))$$

$$= \mathbb{E}_{\boldsymbol{z}\sim p_{\mathcal{N}}} \log p_{\mathcal{N}}(\boldsymbol{z}) - \log p_X(\boldsymbol{x}) - \log \det(\partial_{\boldsymbol{z}} \boldsymbol{x})$$

Recall $p_Z(\boldsymbol{z}) = p_X(\boldsymbol{x}) \det(\partial_{\boldsymbol{z}} \boldsymbol{x})$

Machine Learning Holographic Mapping by Neural Network Renormalization Group

H Hu, S-H Li, L Wang, Y-Z You. arXiv: 1903.00804



Hong-Ye Hu Shuo-Hui Li Lei Wang

(UCSD → Harvard) (IOP, CAS)

Problem Setup

- Given a statistical mechanics/field theory model

$$-\log p(\boldsymbol{x}) = S(\boldsymbol{x})$$

- Train a bijective RG transformation $\boldsymbol{z} = R(\boldsymbol{x})$ to minimize the bulk mutual information (disentangle latent variables $\boldsymbol{z}$)

$$\begin{aligned}\mathcal{L} &= D_{\mathrm{KL}}(p_{\mathcal{N}}(\boldsymbol{z}) \| p_Z(\boldsymbol{z})) \\ &= \mathbb{E}_{\boldsymbol{z}\sim p_{\mathcal{N}}} \log p_{\mathcal{N}}(\boldsymbol{z}) + S(\boldsymbol{x}) - \log\det(\partial_{\boldsymbol{z}}\boldsymbol{x})\end{aligned}$$

- After training
  - Generation task: generate $\boldsymbol{x}$ from $\boldsymbol{z}$

$$p_X(\boldsymbol{x}) = p_{\mathcal{N}}(\boldsymbol{z})\det(\partial_{\boldsymbol{x}}\boldsymbol{z}) \Rightarrow \langle x_i x_j\rangle = \mathbb{E}_{\boldsymbol{x}\sim p_X} x_i x_j$$

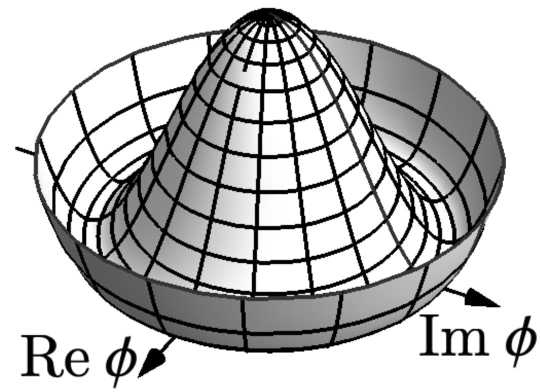
  - Inference task: infer $\boldsymbol{z}$ from $\boldsymbol{x}$

$$S_{\mathrm{eff}}(\boldsymbol{z}) = S(\boldsymbol{x}) - \log\det(\partial_{\boldsymbol{z}}\boldsymbol{x}) \Rightarrow \langle z_i z_j\rangle_{\mathrm{eff}}$$

Complex ϕ^4 Model in 2D

- Lattice field theory on square lattice

$$S[\phi] = -t \sum_{\langle ij \rangle} \phi_i^* \phi_j + \sum_i (\mu |\phi_i|^2 + \lambda |\phi_i|^4)$$



- Symmetry: internal U(1) $\phi_i \to \mathrm{e}^{\mathrm{i}\alpha} \phi_i$
- Effectively 2D XY model $\phi_i = \sqrt{\rho} e^{\mathrm{i}\theta_i}$

$$S[\theta] = -\frac{1}{T} \sum_{\langle ij \rangle} \cos(\theta_i - \theta_j)$$

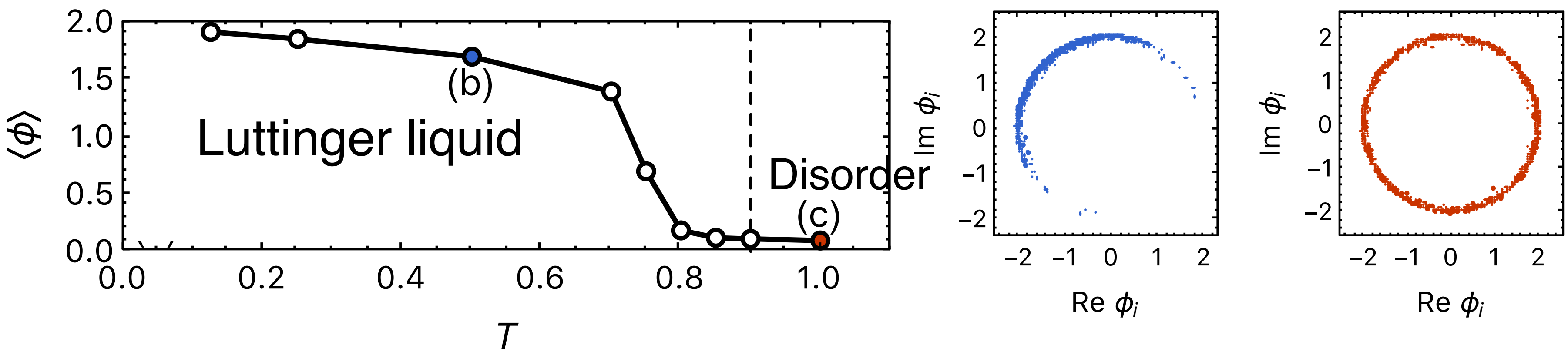
$\langle \phi_i^* \phi_j \rangle \sim r_{ij}^{\alpha}$ $\langle \phi_i^* \phi_j \rangle \sim e^{-r_{ij}/\xi}$

T (effectively tuned by t)

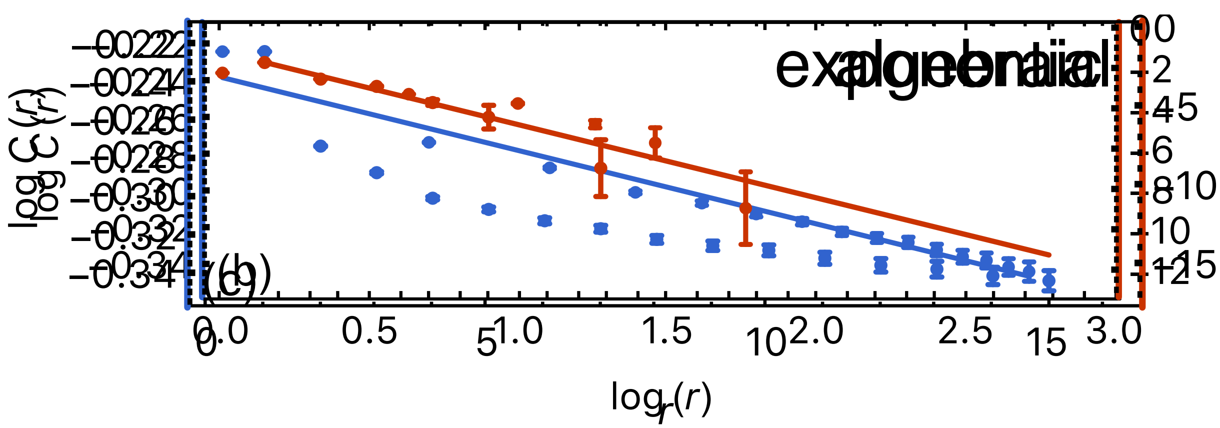
Luttinger liquid (CFT) — T_{KT} — Disordered

Performance of the Generative Model

- Let us first make sure that the machine learns the correct physics from the given action.
  - Phase diagram (32x32 finite size lattice)



  - Correlation function $\langle \phi_i^* \phi_j \rangle = \mathbb{E}_{\boldsymbol{\phi}\sim\text{flow}} \phi_i^* \phi_j$



Probing Holographic Bulk Geometry

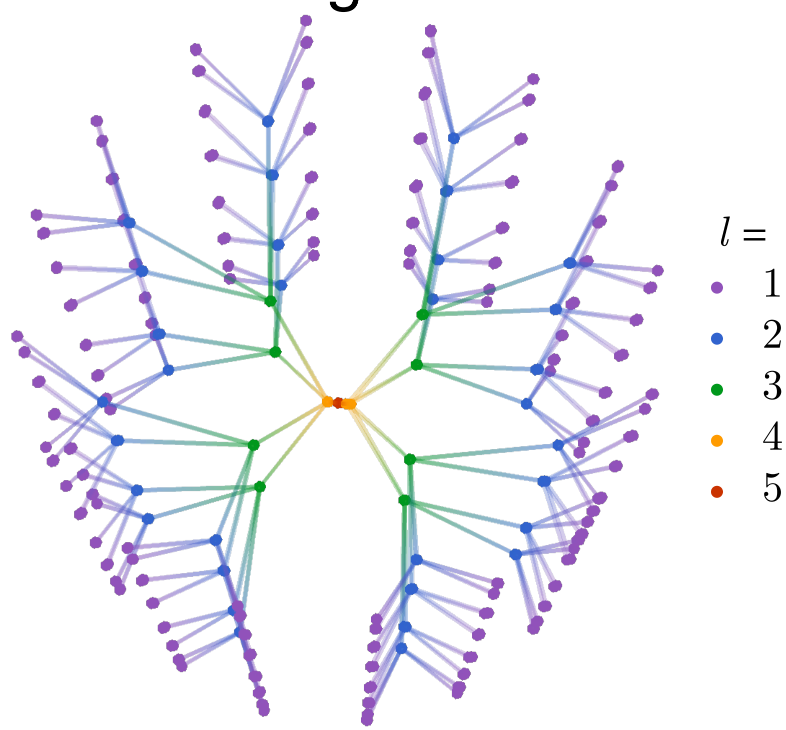
- We can further infer the bulk effective action (the holographic dual theory)

$$S_{\mathrm{eff}}(\boldsymbol{z}) = S(\boldsymbol{x}) - \log\det(\partial_{\boldsymbol{z}}\boldsymbol{x})$$

and evaluate the bulk correlation $\langle z_i z_j \rangle_{\mathrm{eff}}$, which is expected to tell us about the bulk geometry, as the bulk geodesic distance can be inferred from

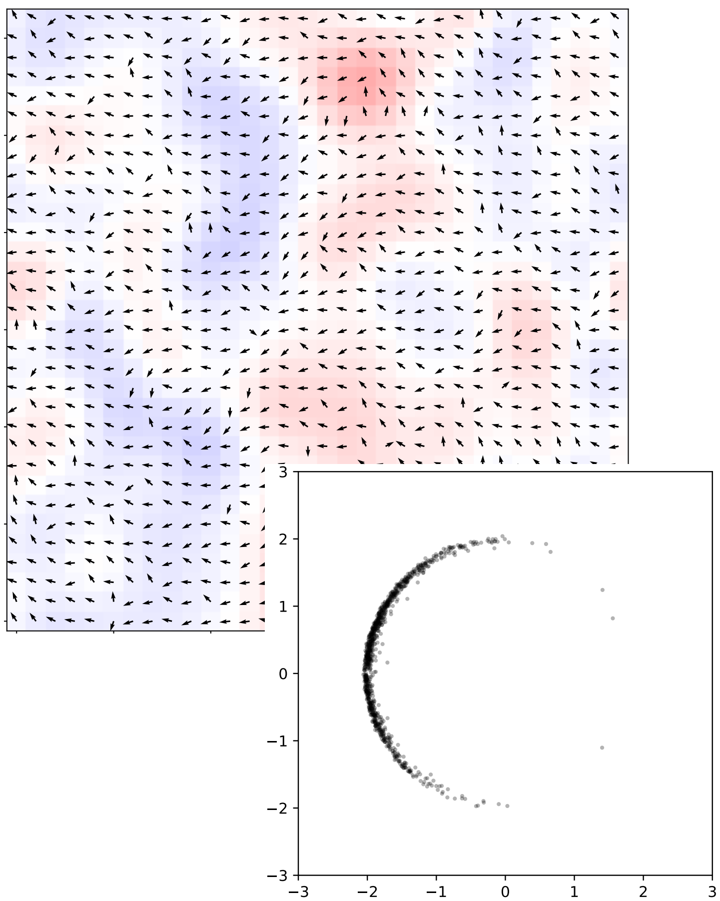
$$d_{ij} = d_0 - \xi \ln\langle z_i z_j \rangle_{\mathrm{eff}}$$

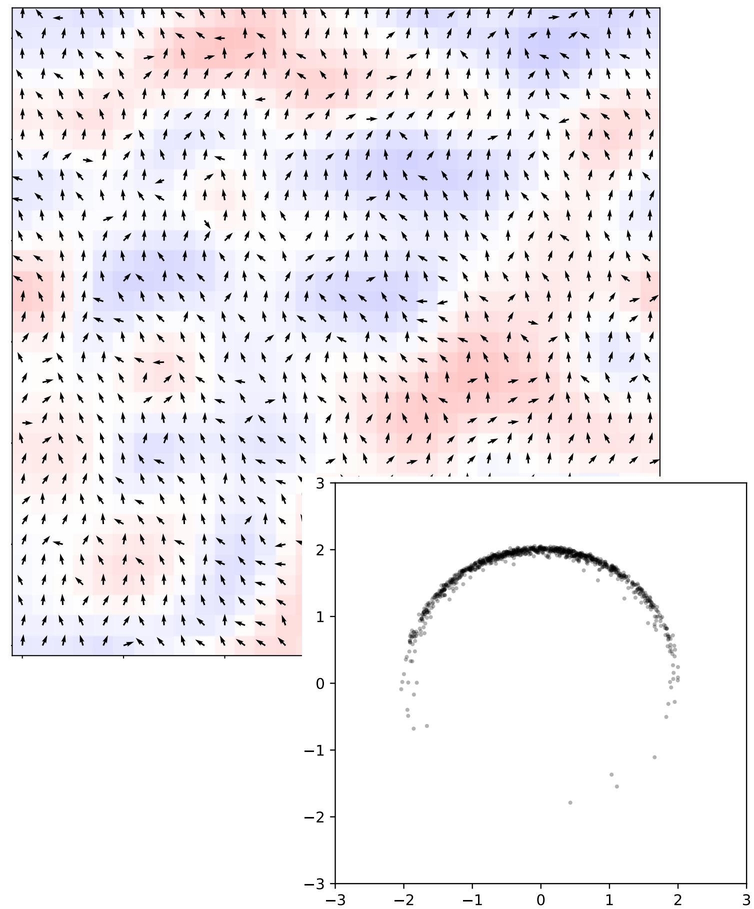
- In the CFT phase, the distance matches hyperbolic geometry (~ AdS), verifying the AdS/CFT correspondance!



Efficient Sampling from the Bulk

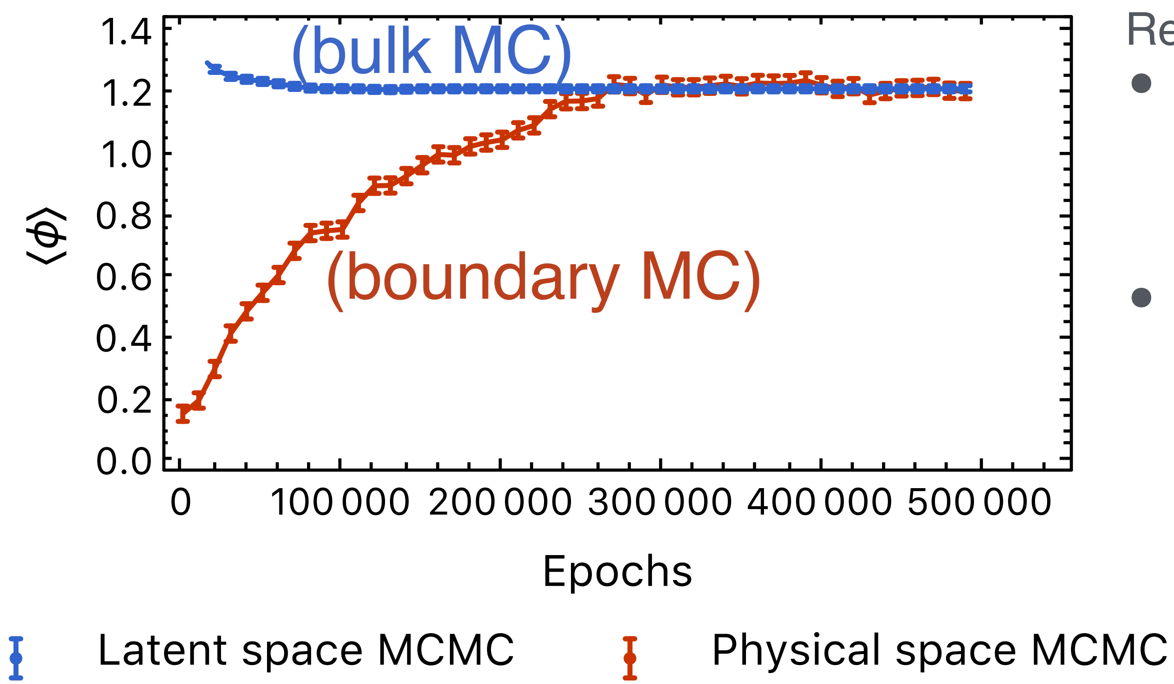
- Sampling: holographic mapping from bulk to boundary
  - Massive field in the bulk → Critical field on the boundary
  - Local update in the bulk → Global update on the boundary





Efficient Sampling from the Bulk

- Sampling: holographic mapping from bulk to boundary
  - Massive field in the bulk → Critical field on the boundary
  - Local update in the bulk → Global update on the boundary
- Order parameters converges faster using bulk MCMC.



Related topics:

- Self-learning MC
  Huang, Wang, PRB (2017)
  Liu, Qi, Meng, Fu, PRB(2017)
  …
- Super-resolution sampling
  Efthymiou, Beach, Melko (2019)

Neural Network Renormalization Group Applied to Computer Vision

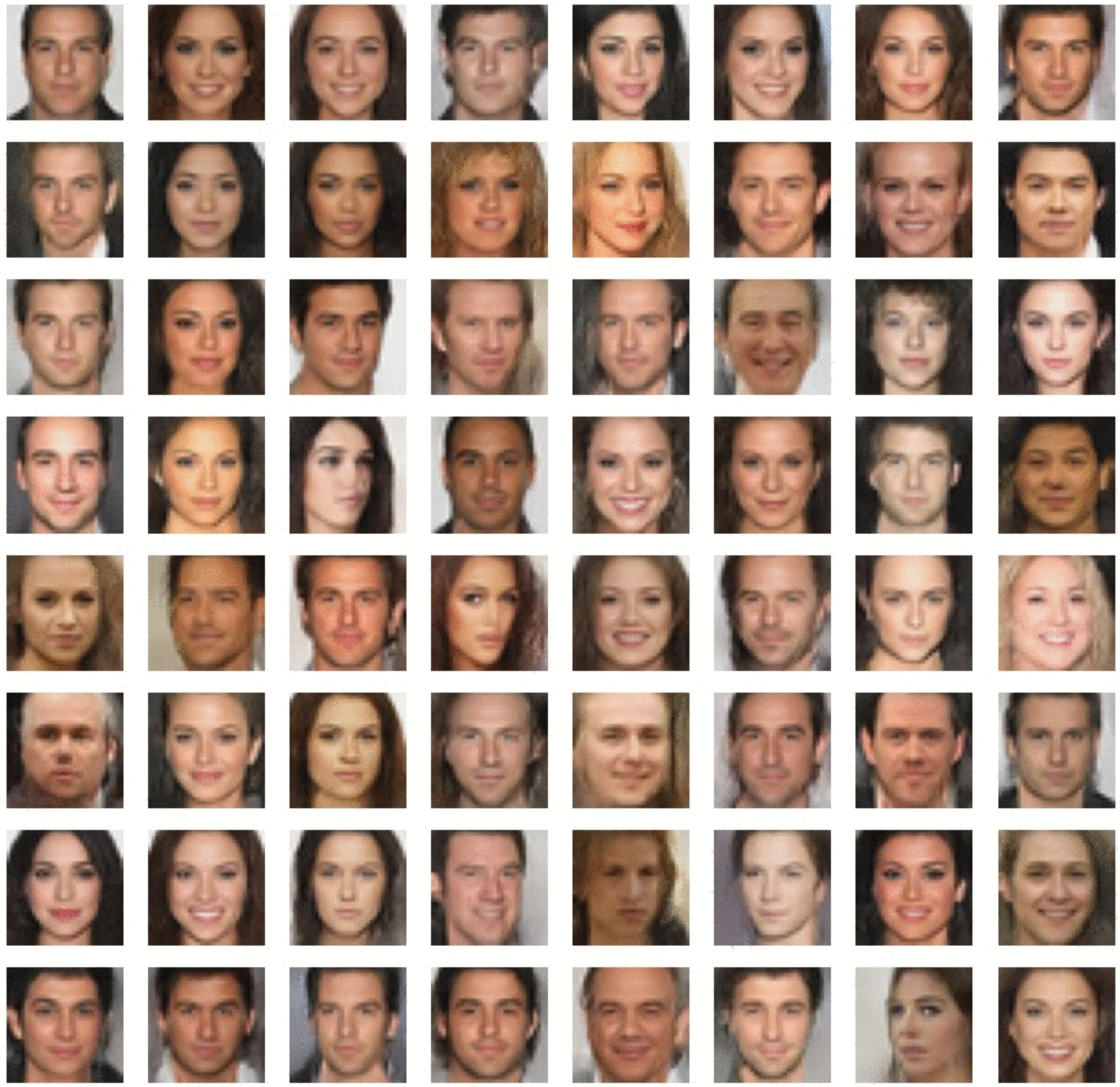
H Hu, D Wu, Y-Z You, B Olshausen, Y Chen. arXiv: 2010.00029




Hong-Ye Hu (UCSD → Harvard)
Dian Wu (EPFL)
Yubei Chen, Bruno Olshausen (UC Berkeley)

Human Face Dataset (CelebA)

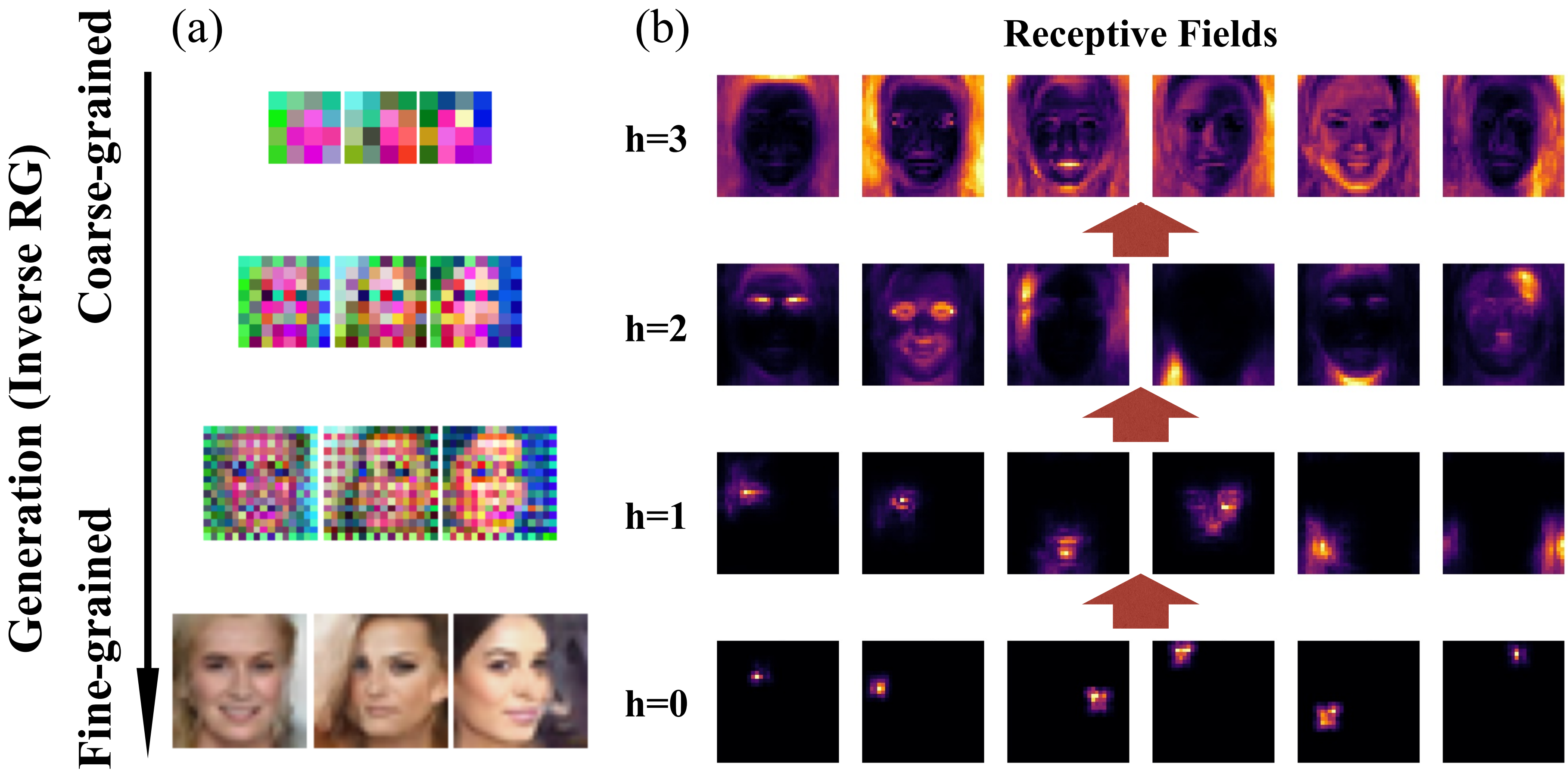
- Samples generated by RG-Flow:



Emergent Hierarchical Representations

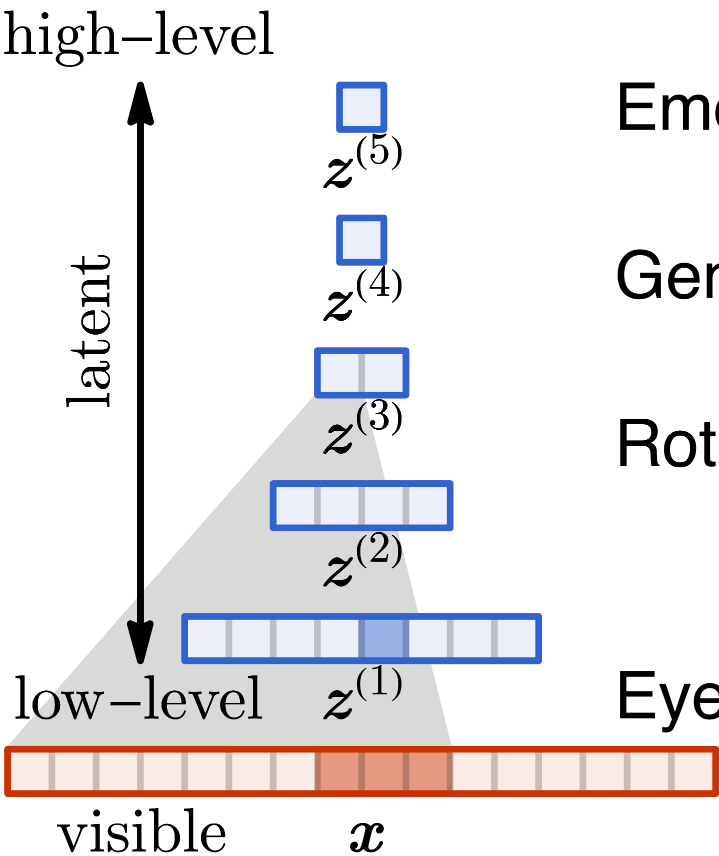
- After training, probe how visible variables respond to perturbations of latent variables.

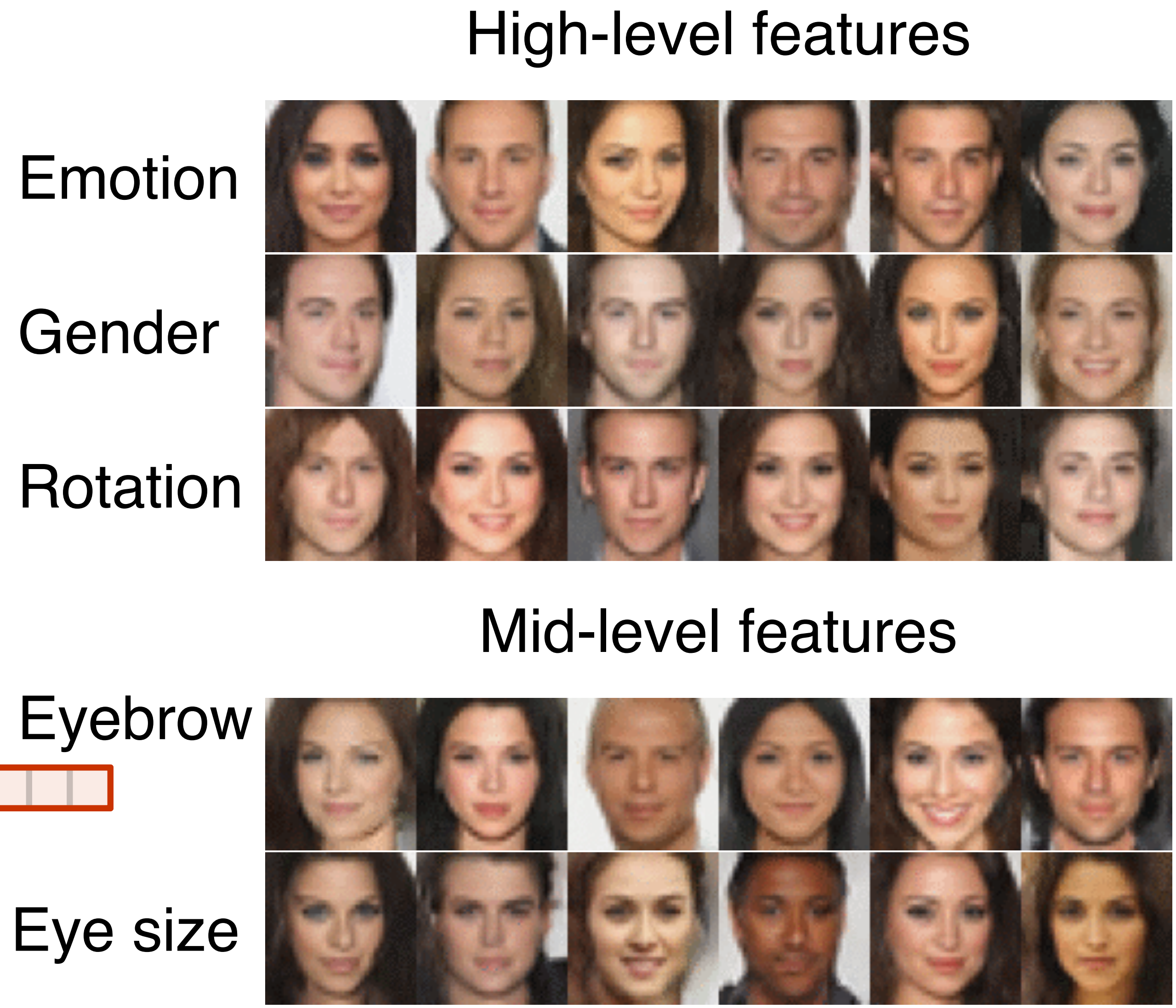
Emergent features (unsupervised)



Emergent Hierarchical Representations

- Different latent variables control features at different scales

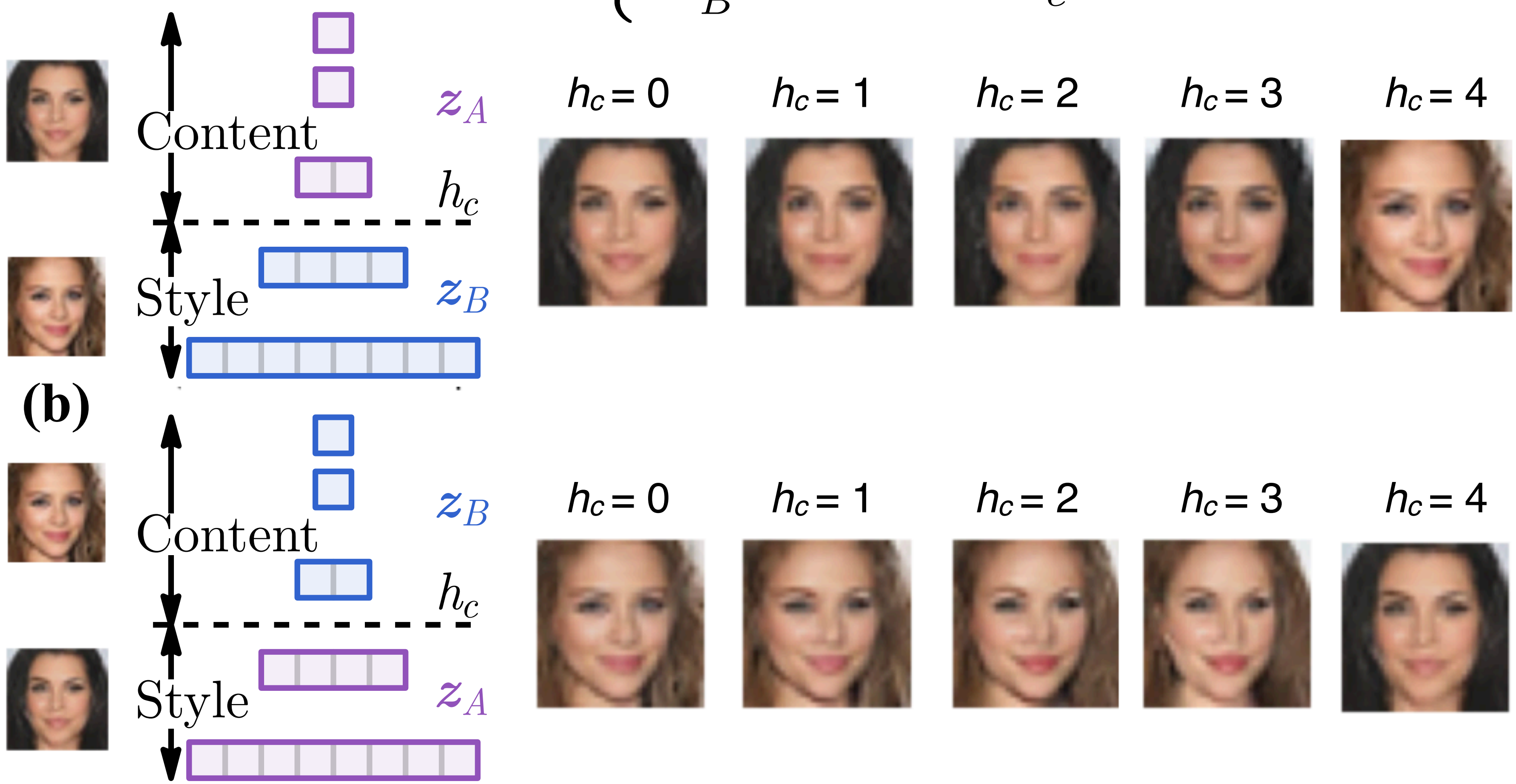




Multi-Scale Feature Mixing

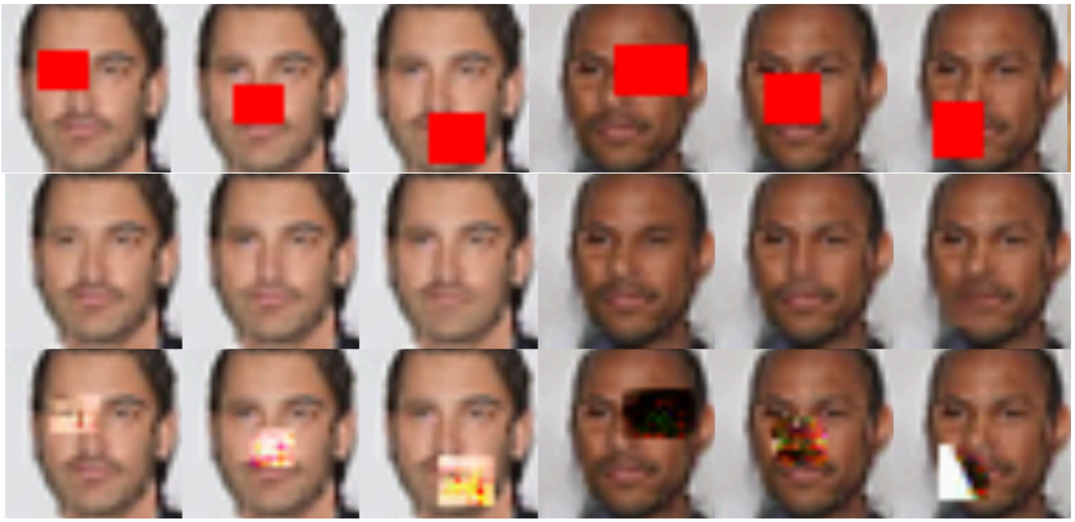
- Mixing high-level features of A with low-level features of B

$$z^{(h)} = \begin{cases} z_A^{(h)} & \text{for } h \geq h_c, \\ z_B^{(h)} & \text{for } h < h_c \end{cases}$$



Efficient Error Correction

- Error correction: restore locally corrupted images

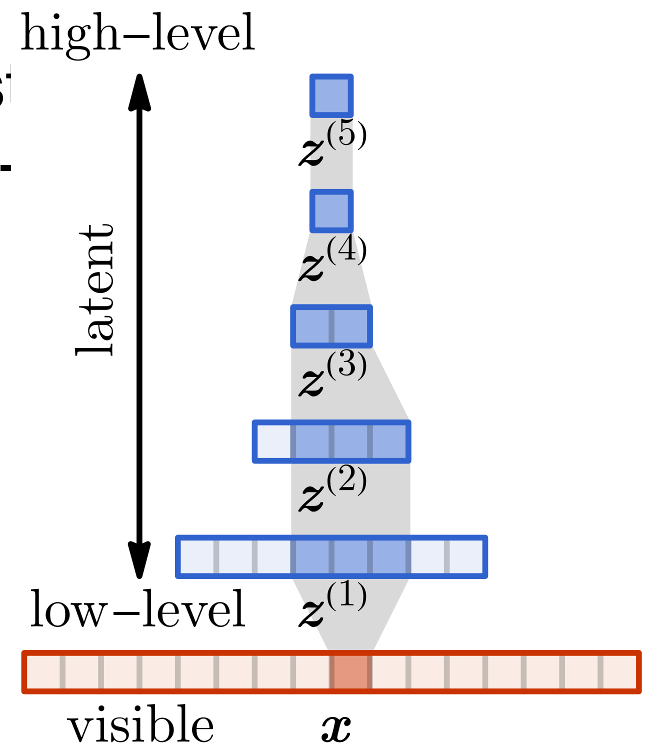


Corrupted image

RG-Flow (ours)

Rest
real-

- Light-cone volume ~ O(log N) in the hyperbolic space
- Only resample ~ O(log N) latent variables for error correction (more efficient than O(N) scaling)



Summary

- RG-Flow: a hierarchical *flow-based generative model* motivated by the *renormalization group* in physics.
  - ML helps to find optimal RG scheme, holographic latent variables, and “gravitational-dual” theory …
  - Holographic duality helps boost sampling efficiency, multi-scale feature tuning/mixing, error correction …
- Applicable to:
  - Quantum field theory/statistical physics.
  - Image processing processing.

Thanks for your attention!