Learning Hierarchical Features with Joint Latent Space Energy-Based Prior

Jiali Cui[1] Ying Nian Wu[2] Tian Han[1]

[1]Stevens Institute of Technology
[2]University of California, Los Angeles (UCLA)

September 21, 2023

Multi-layer Generator Model

- Generator Model can be specified using joint distribution:

$$p_\beta(\mathbf{x}, \mathbf{z}) = p_{\beta_0}(\mathbf{x}|\mathbf{z})p_{\beta_{>0}}(\mathbf{z})$$

- Multi-layer Generator Model consists of multiple layers of latent variables:

$$p_{\beta_{>0}}(\mathbf{z}) = \prod_{i=1}^{L-1} p_{\beta_i}(\mathbf{z}_i|\mathbf{z}_{i+1})p(\mathbf{z}_L)$$

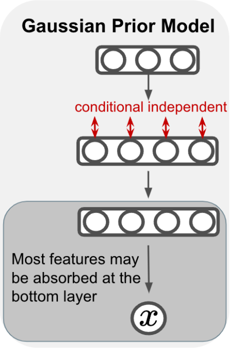
- Conditional Gaussian Distribution

$$p_{\beta_i}(\mathbf{z}_i|\mathbf{z}_{i+1}) \sim \mathcal{N}(\mu_{\beta_i}(\mathbf{z}_{i+1}), V_{\beta_i}(\mathbf{z}_{i+1}))$$

Hierarchical Representation Learning

Limitation

1. Such multi-layer generator models can learn most data representation at the bottom layer.
2. Intra-layer relation among latent units is ignored.
3. conditional independent



Hierarchical Representation Learning

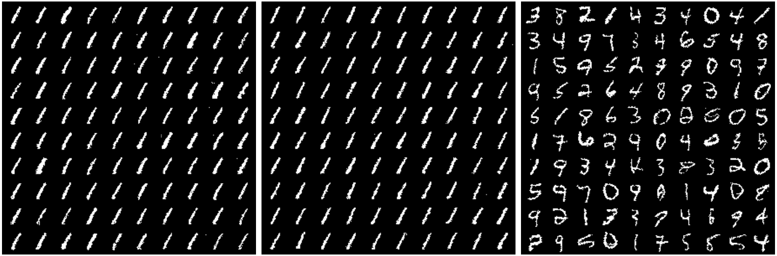


Figure: Hierarchical Sampling.

Latent Space Energy-based Model

- Latent space energy-based model (EBM) can be specified as:

$$p_\alpha(\mathbf{z}) = \frac{1}{\mathbf{Z}(\alpha)} \exp[f_\alpha(\mathbf{z})] p_0(\mathbf{z})$$

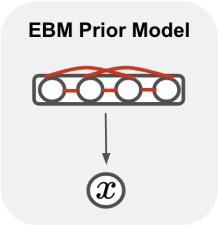
where $\mathbf{Z}(\alpha)$ is the normalizing function, $f_\alpha(\mathbf{z})$ is the energy function, and $p_0(\mathbf{z})$ is the referenced distribution.

Hierarchical Representation Learning

Limitation

1. Single-layer latent space.
2. Mixed representation learned.

non-hierarchical structure



Hierarchical Representation Learning

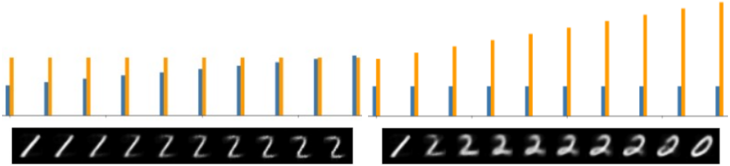


Figure: Latent traverse.

Joint Latent Space EBM

- Joint Latent Space EBM Prior

$$p_\alpha(\mathbf{z}) = \frac{1}{\mathbf{Z}(\alpha)} \exp[f_\alpha([\mathbf{z}_1, \ldots, \mathbf{z}_L])] p_0([\mathbf{z}_1, \ldots, \mathbf{z}_L])$$

- Architectural Generation Model

$$\begin{aligned} &h_L = g_L(\mathbf{z}_L) \\ &h_i = g_i([\mathbf{z}_i, h_{i+1}]), \quad i = 1, 2, \ldots, L-1 \\ &\mathbf{x} \sim \mathcal{N}(h_1, \sigma^2 I_D) \end{aligned}$$

Joint Latent Space EBM

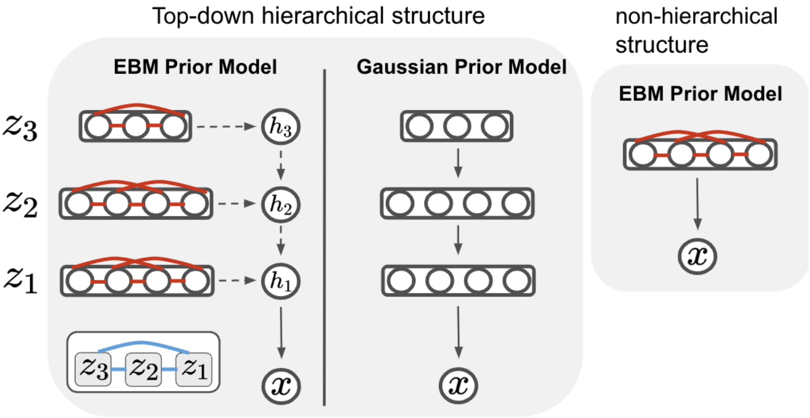


Figure: Illustration.

Hierarchical Representation Learning

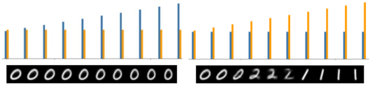


Figure: Latent traverse.

Figure: Hierarchical sampling on MNIST.

Figure: Hierarchical sampling on SVHN.

Expressivity

Method	SVHN	CelebA-64
ABP	49.71	51.50
LVAE	39.26	53.40
BIVA	31.65	33.58
SRI	35.23	36.84
VLAE	43.95	44.05
2s-VAE	42.81	44.40
RAE	40.02	40.95
NCP-VAE	33.23	42.07
Multi-NCP	26.19	35.38
LEBM	29.44	37.87
Ours	24.16	32.15

Table: IS(↑) and FID(↓) on CIFAR-10.

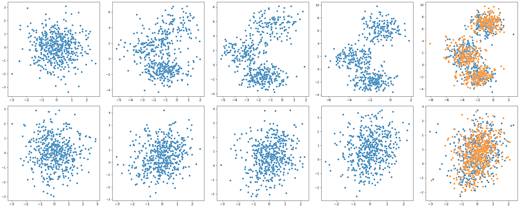


Figure: Visualization of Latent Space.

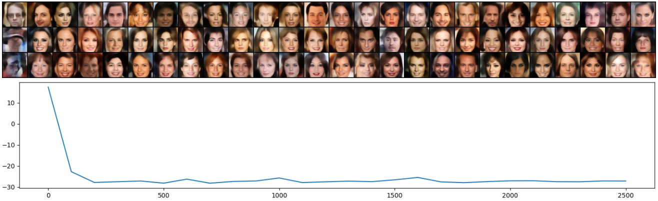


Figure: Long-run Langevin traverse.