Learning Joint Latent Space EBM Prior Model for Multi-layer Generator

Jiali Cui[1] Ying Nian Wu[2] Tian Han[1]
[1]Stevens Institute of Technology [2]University of California, Los Angeles

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Problem

Multi-layer Generator Model: For the multi-layer generator model, the prior model is hierarchical and can be specified as

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

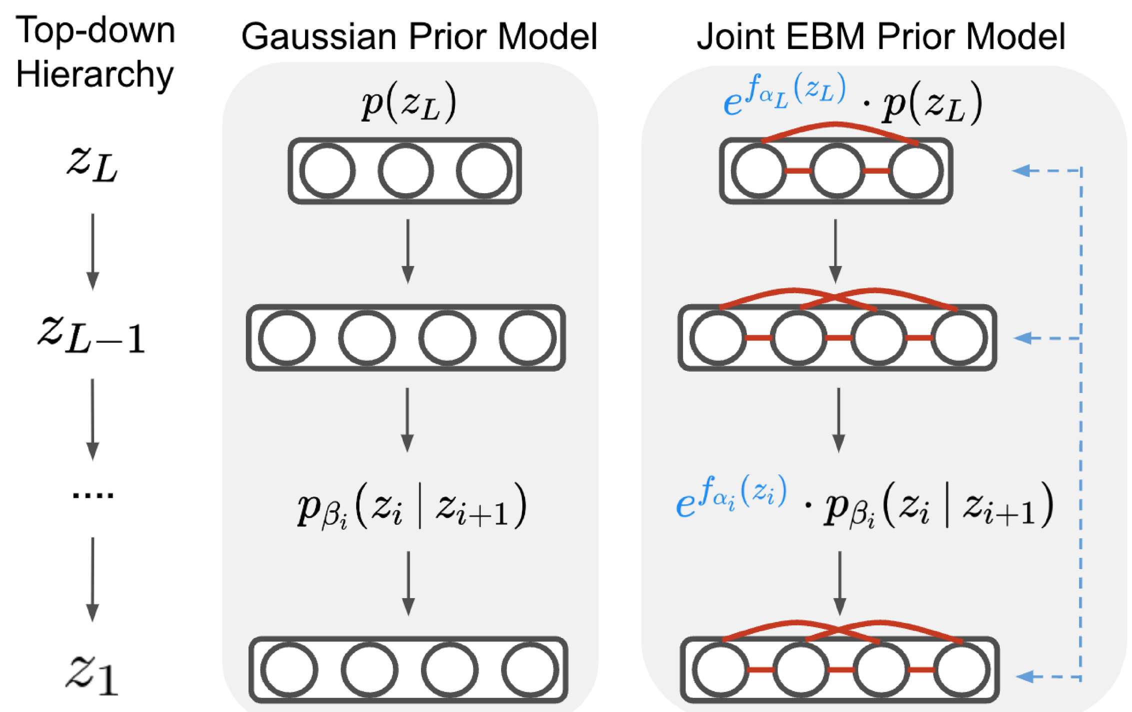
Limitation: Such a prior model focused on *inter-layer* modeling while ignoring the *intra-layer* contextual modeling as the latent units are *conditional independent* within each layer.

Methodology

Joint Latent Space EBM Prior Model: We propose the joint EBM prior for multi-layer generator models, which can effectively capture the *intra-layer* relations at each layer and jointly correct the latent variables from all layers.

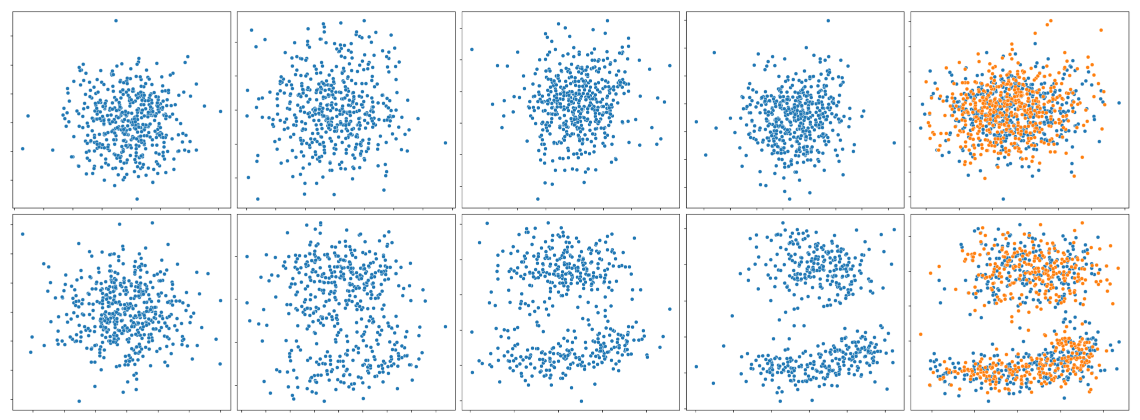
$$p_{\alpha,\beta_{>0}}(\mathbf{z}) = \frac{1}{Z_{\alpha,\beta_{>0}}} \exp\left[\sum_{i=1}^{L} f_{\alpha_i}(\mathbf{z}_i)\right] \prod_{i=1}^{L-1} p_{\beta_i}(\mathbf{z}_i|\mathbf{z}_{i+1})p(\mathbf{z}_L)$$

Comparison with Gaussian Prior Model:



Black solid lines with arrow: inter-layer relations modelling. **Red solid lines**: intra-layer contextual relations modelling. **Blue dashed lines:** joint modelling upon all layers.

Toy MNIST with only '0' and '1' digits available.



Langevin transition on latent codes (bottom: $\mathbf{z}_1$, top: $\mathbf{z}_2$). **Blue, Orange** color indicate prior and posterior, respectively. We use 2-dimensional latent codes and show the transition of Langevin dynamics on each layer, where the Gaussian prior can be successfully tilted via EBM to match the multi-modal posterior.

Experiment: Image Synthesis

Image synthesis on CelebA-HQ-256

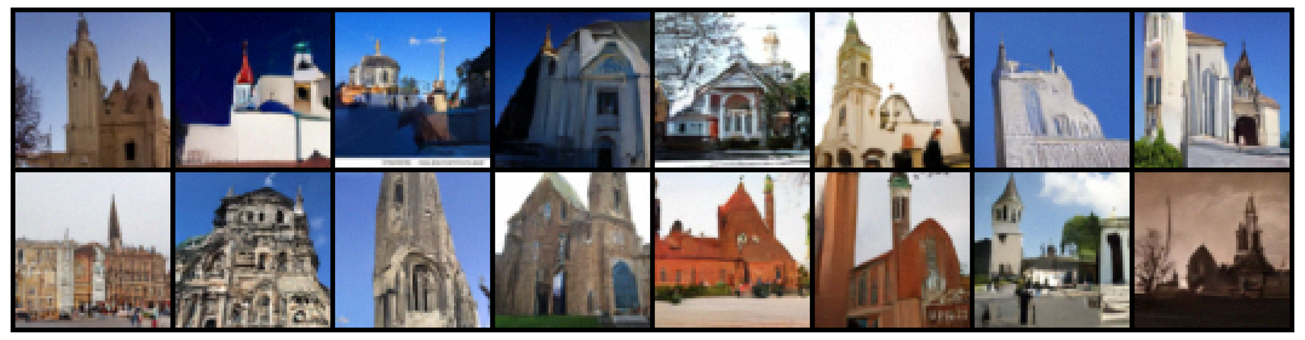
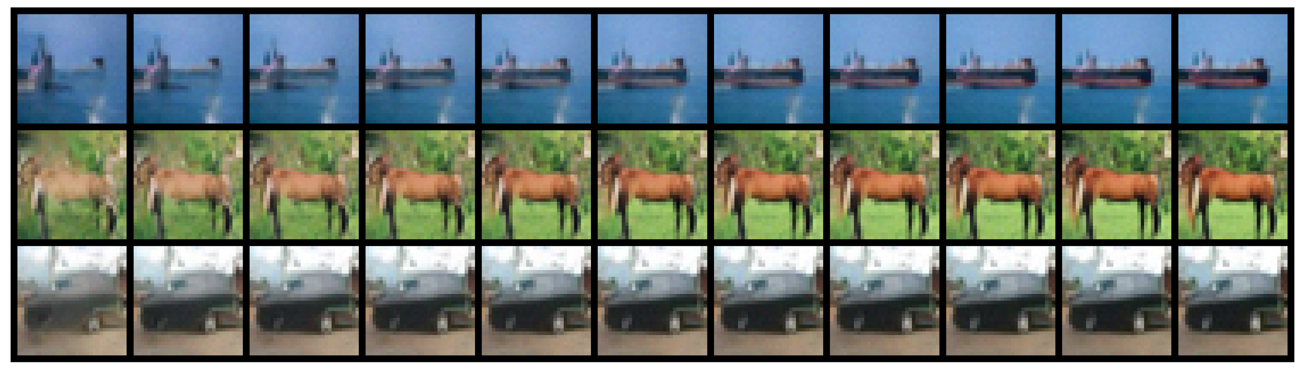


Image synthesis on LSUN-Church-64



Langevin transition on CIFAR-10

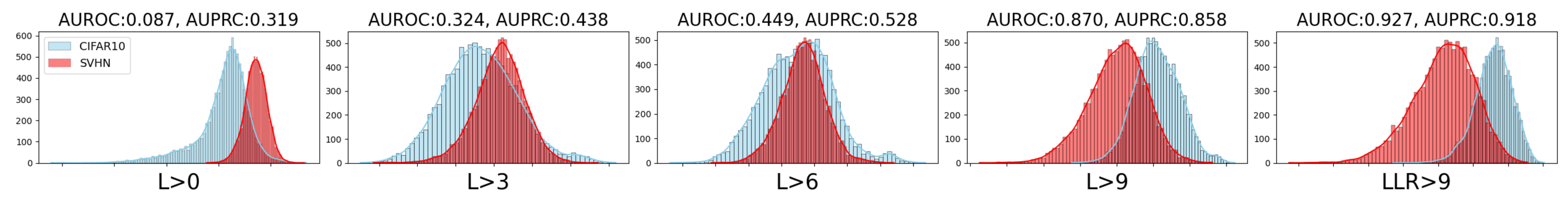
CIFAR-10	IS	FID
NVAE*	5.30	37.73
Ours	**8.99**	**11.34**
NCP-VAE	-	24.08
VAEBM	8.43	12.19
Other EBMs		
IGEBM	6.78	38.2
ImprovedCD	7.85	25.1
Divergence Triangle	-	30.10
Adv-EBM	9.10	13.21
Other Likelihood Models		
GLOW	3.92	48.9
PixelCNN	4.60	65.93
GANs+Score-based Models		
BigGAN	9.22	14.73
StyleGANv2 w/o ADA	8.99	9.9
NCSN	8.87	25.32
DDPM	9.46	3.17

Experiment: Hierarchical Representations

Out-of-distribution Detection: For our joint EBM, we compute the adapted unnormalized decision function as

$$LLR^{>k}_{\text{EBM}} = L^{>0}_{\text{EBM}} - L^{>k}_{\text{EBM}}$$

$$L^{>k}_{\text{EBM}} = \mathbb{E}_{\mathbf{z}_{>k}\sim q_\omega(\mathbf{z}|\mathbf{x}), \mathbf{z}_{\le k}\sim p_{\beta_{>0},\alpha}(\mathbf{z})}\left[\log p_{\beta_0}(\mathbf{x}|\mathbf{z}) + \log p_{\beta_{>0}}(\mathbf{z}) + \sum_{i=1}^{L} f_{\alpha_i}(\mathbf{z}_i)\right]$$



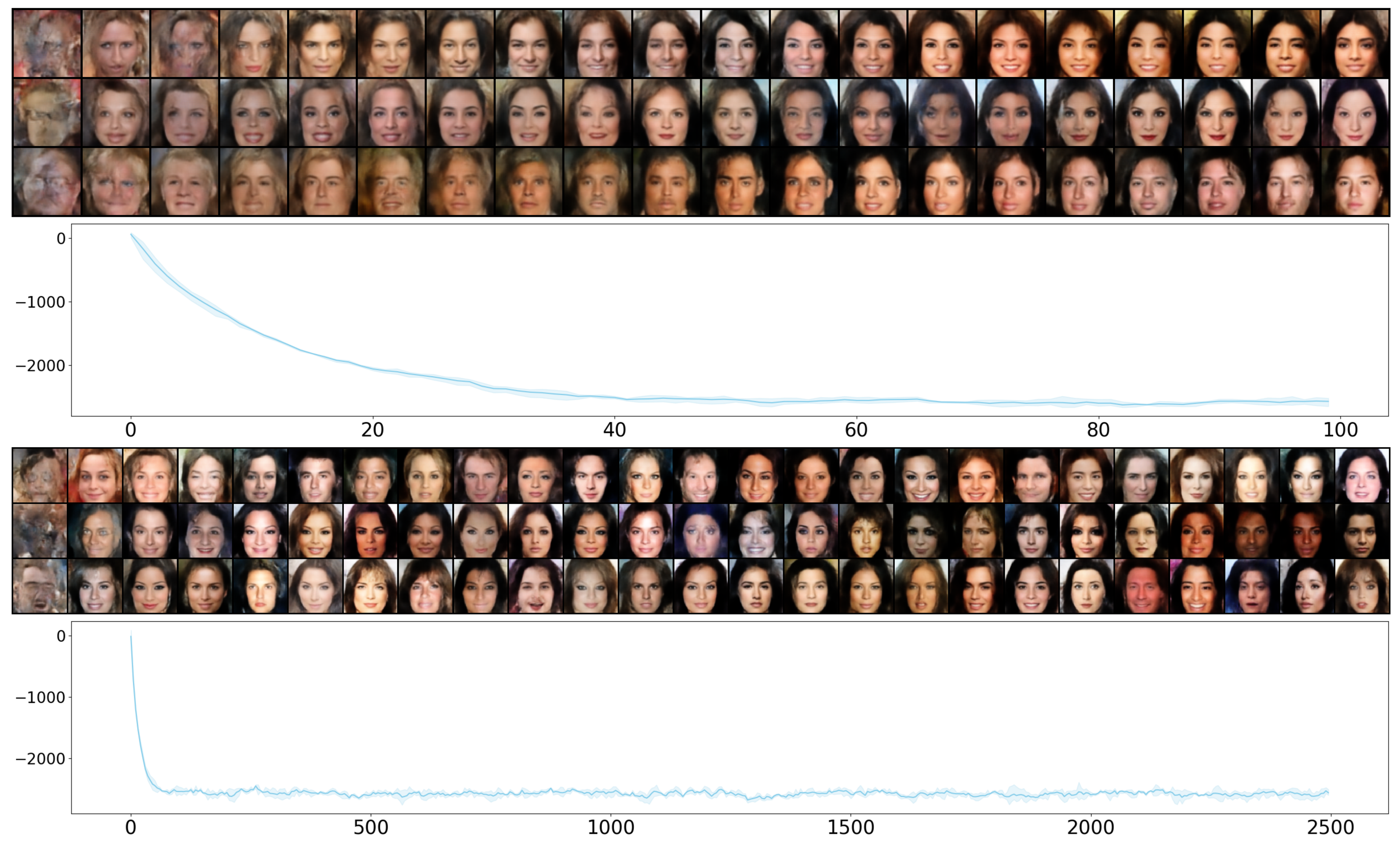
Histograms of density of $L^{>k}_{\text{EBM}}$ for CIFAR-10 (in) / SVHN (out).

Hierarchical Sampling:

Hierarchical sampling for Gaussian prior model (**bottom**) and EBM prior model (**top**). From **left panel** to **right panel**, latent vectors are sampled from the bottom layers to the top layers.

Experiment: Analysis of Latent Space

Long-run langevin transition:



Trajectory in data space and energy profile. **Top:** Langevin transition with 100 steps. **Bottom:** Langevin transition with 2500 steps.

Anomaly Detection: MNIST with one digit of data being held out as anomaly for training, and both normal (e.g., other nine digits) and anomalous data are used for testing.

Heldout Digit	1	4	5	7	9
VAE	0.063	0.337	0.325	0.148	0.104
MEG	0.281 ± 0.035	0.401 ± 0.061	0.402 ± 0.062	0.290 ± 0.040	0.342 ± 0.034
BiGAN-σ	0.287 ± 0.023	0.443 ± 0.029	0.514 ± 0.029	0.347 ± 0.017	0.307 ± 0.028
OT-SRI	0.353 ± 0.021	0.770 ± 0.024	0.726 ± 0.030	0.550 ± 0.013	0.555 ± 0.023
LEBM	0.336 ± 0.008	0.630 ± 0.017	0.619 ± 0.013	0.463 ± 0.009	0.413 ± 0.010
Ours	$\mathbf{0.470 \pm 0.009}$	$\mathbf{0.941 \pm 0.001}$	$\mathbf{0.964 \pm 0.003}$	$\mathbf{0.815 \pm 0.004}$	$\mathbf{0.796 \pm 0.004}$

AUPRC scores for unsupervised anomaly detection where we use un-normalized log-posterior $L^{>0}_{\text{EBM}}$ as our decision function