Supplementary Material for Learning Hierarchical Features with Joint Latent Space Energy-Based Prior

Jiali Cui[1], Ying Nian Wu[2], Tian Han[1]
[1]Department of Computer Science, Stevens Institute of Technology
[2]Department of Statistics, University of California, Los Angeles
{jcui7,than6}@stevens.edu, ywu@stat.ucla.edu

1. Additional Experiment

In order to better understand the proposed method, we conduct additional experiments, including (i) out-of-distribution detection and (ii) disentanglement learning, which further explore the potential of our method in various challenging tasks.

1.1. Out-of-Distribution Detection

We conduct the out-of-distribution (OOD) detection for our model. We train our model on Fashion-MNIST (in-distribution) with MNIST (OOD) being the test dataset. We consider baseline models, including VLAE and HVAE, in which HVAE belongs to the *conditional hierarchical model* (see Sec. 2.1), allowing a specialized decision function (likelihood-ratio, LLR [3]) to be applied. It can be hard to directly apply such decision function to the our *architectural hierarchical model*, thus we use the unnormalized log-posterior as the decision function (same as Anomaly Detection in Sec. 5.4) for our model and report the results in Tab. 1. It can be seen that our model shows superior performance compared to VLAE, while HVAE could render better performance by using the specialized LLR decision function.

Table 1. Fashion-MNIST (in) vs MNIST (out)

	HVAE(LLR$^{>k}$)	HVAE	VLAE	**Ours**
AUPRC ↑	0.984	0.363	0.344.	0.893
AUROC ↑	0.984	0.268	0.199	0.897
#param	3.4M	3.4M	3.2M	3.2M

1.2. Disentanglement Learning

We examine the disentanglement learning of our model. We train our model on 3DShapes dataset[1] with $L = 3$ layers and visualize the traverse of each latent dimension in Figure 1. We could see that the semantic factors, such as shape, size and direction, are disentangled into the latent dimensions of the top layer (i.e., $\mathbf{z}_3$), while low-level factors, such as background and object color, are disentangled into the latent dimensions of lower layers(i.e., $\mathbf{z}_1$, $\mathbf{z}_2$).

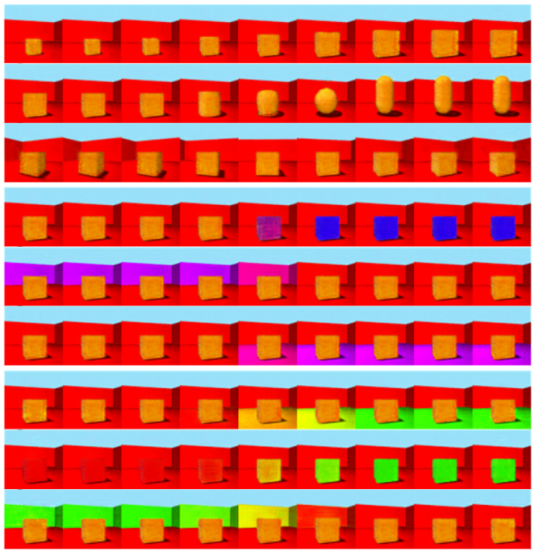

Figure 1. Disentanglement traverse of each latent unit of our model. **Top panel (top 3 rows):** the top layer $\mathbf{z}_3$, and $\mathbf{z}_{3_{dim}} = 3$. The top row of each panel illustrates the traverse on the first unit.

Table 2. MIG and MIG-sup on 3DShapes. s denotes the progressive steps of pro-VLAE.

	Ours	pro-VLAE $s = 3$	pro-VLAE $s = 2$	pro-VLAE $s = 1$
MIG	**0.554**	0.357	0.339	0.247
MIG-sup	**0.672**	0.406	0.333	0.136

We further quantitatively evaluate our model. Prior work [4] applies the progressive learning strategy on VLAE to improve disentangled factor learning and computes MIG[2] and MIG-sup[4] as the measurement. The pro-VLAE with $\beta = 1$ (without progressive learning) is then considered as our baseline model. We train our model with the same inference and generator model and use the same latent dimension as the baseline model, and we compare with the pro-VLAE that uses multiple progressive steps (i.e., steps = 1, 2, 3). The numbers of pro-VLAE are obtained by the code[1], and the comparison is shown in Table 2.

[1]https://github.com/Zhiyuan1991/proVLAE

References

[1] Chris Burgess and Hyunjik Kim. 3d shapes dataset. https://github.com/deepmind/3dshapes-dataset/, 2018. 1

[2] Ricky T. Q. Chen, Xuechen Li, Roger B Grosse, and David K Duvenaud. Isolating sources of disentanglement in variational autoencoders. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. 1

[3] Jakob D Drachmann Havtorn, Jes Frellsen, Soren Hauberg, and Lars Maaløe. Hierarchical vaes know what they don't know. In *International Conference on Machine Learning*, pages 4117–4128. PMLR, 2021. 1

[4] Zhiyuan Li, Jaideep Vitthal Murkute, Prashnna Kumar Gyawali, and Linwei Wang. Progressive learning and disentanglement of hierarchical representations. *arXiv preprint arXiv:2002.10549*, 2020. 1