Implement Multi-Scale group on Deep Hierarchical VAE & Analyze different architecture

IOC歐亭昀 李詰琳 謝文棋

1 Abstract

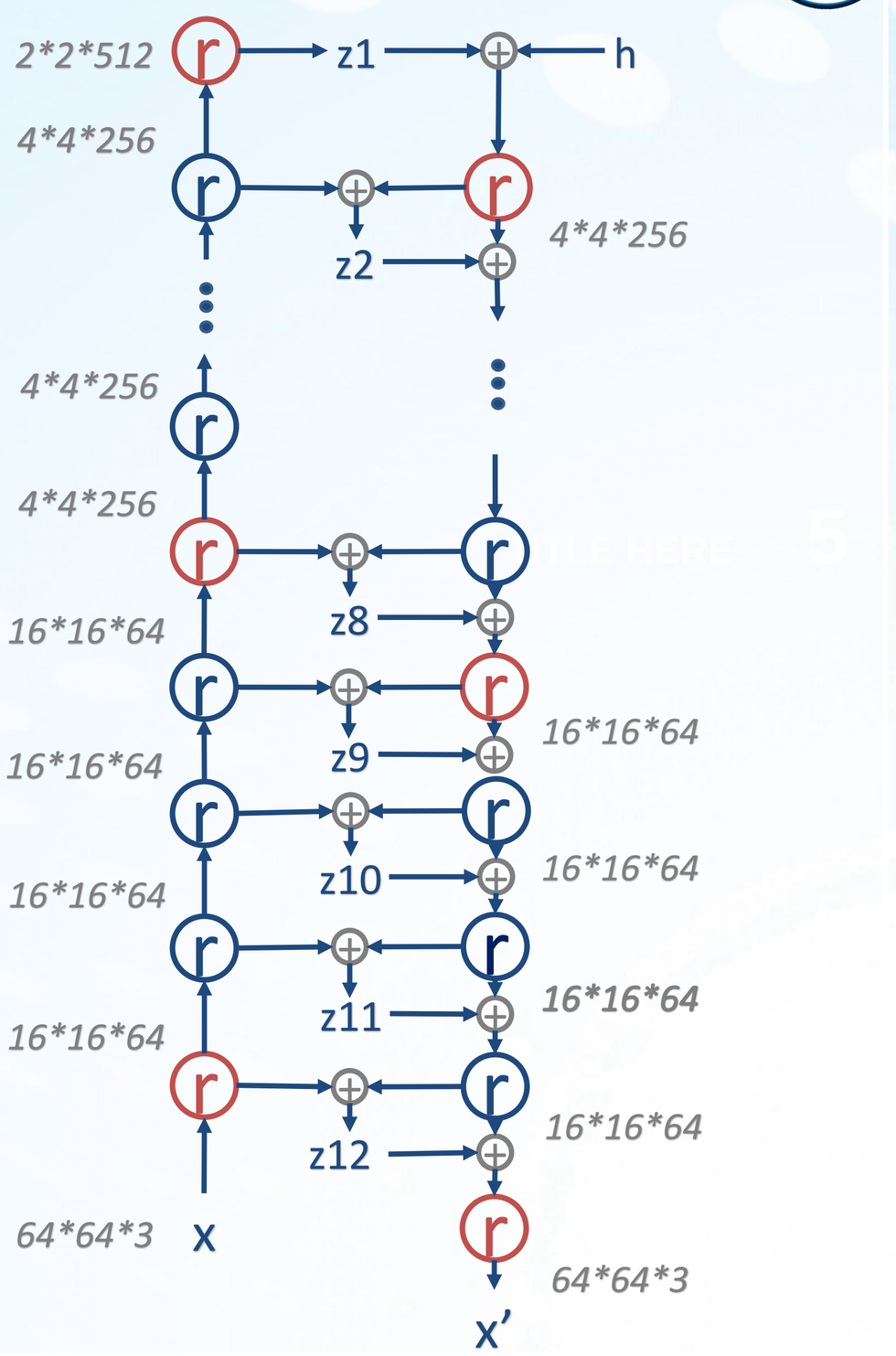
By reading the paper <NVAE: A Deep Hierarchical Variational Autoencoder>, we start to aim at

1.Designing a lightweight Hierarchical VAE model to approximate and reproduce the result on datasets with different complexity, including CelebA(One species) and ImageNet(Various Multiple species).

2.Implementing another improved Multi-Scale group model, which is partitioned into 12 disjoint group of latent variables and three different scale to achieve higher score on both dataset.

3.Futhermore, investigating Variable Control, that is, how the latent variable's each dimension affect the output image's charactxeristic.

2 Architecture



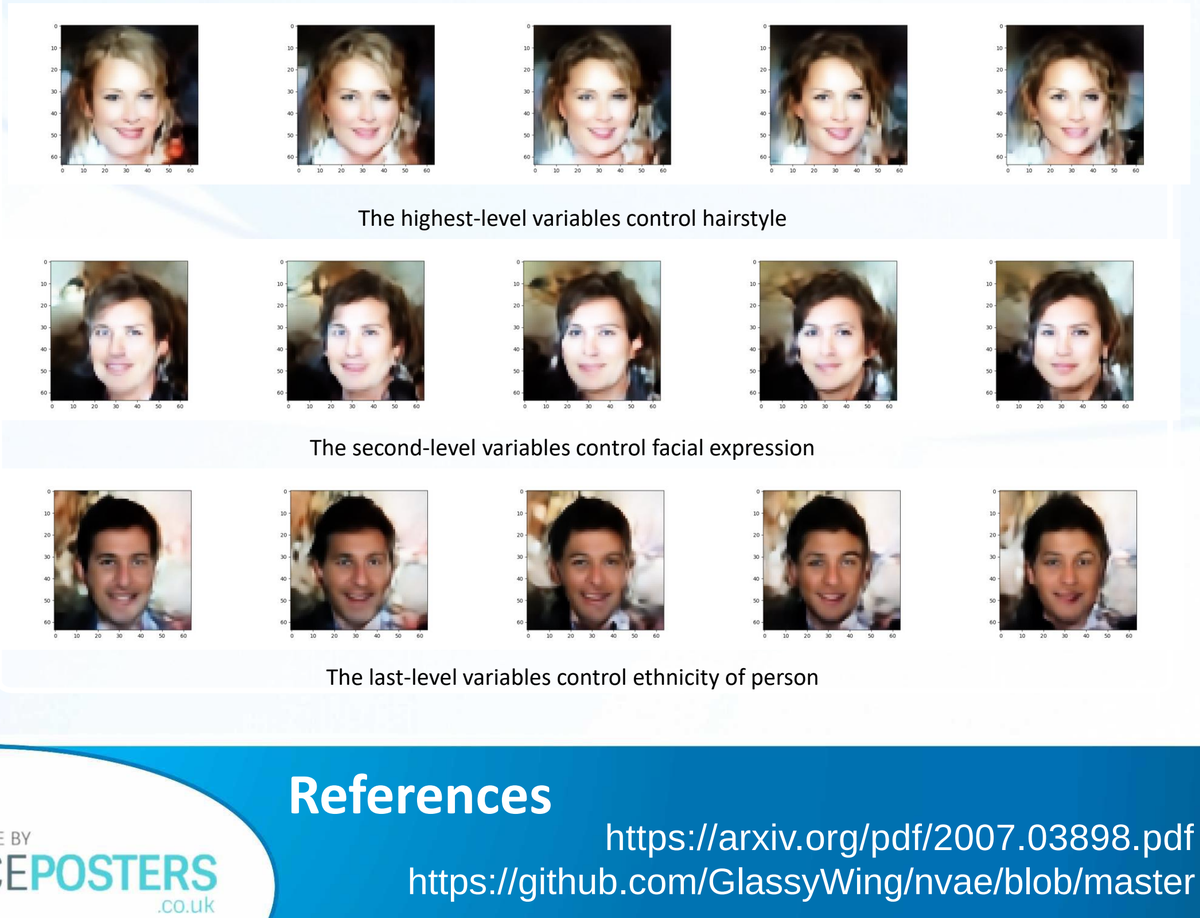
3 CelebA

4 ImageNet

We believe that those two kinds of dataset are with different training goal and complexity. At first CelebA is a dataset with only one species and similar background, so both our two models can converge more quickly and learn much more species detail. On the other hand, ImageNet is with 1000 species and even more complicated background in each species, thus the multi-scale group version model can improve the result more apparently on ImageNet than CelebA.

Different Archi when converge loss to 0.xx(epoch)	CelebA (to 0.70)	ImageNet (to 0.78)
Simple Hierarchical	127	173
Multi-Scale Group	**108**	**146**

5 VariableControl



The highest-level variables control hairstyle

The second-level variables control facial expression

The last-level variables control ethnicity of person

References

https://arxiv.org/pdf/2007.03898.pdf
https://github.com/GlassyWing/nvae/blob/master