Appendix for "Symmetric Variational Autoencoder and Connections to Adversarial Learning"

1 Model Architectures

Table 1: Architecture of the models for sVAE-r on MNIST. BN denotes batch normalization.

Encoder X to z	Decoder z to X	Discriminator
Input 28×28 Gray Image	Input latent code z	Input two 28×28 Gray Image
5×5 conv. 16 ReLU, stride 2, BN	MLP output 1024, BN	5×5 conv. 32 ReLU, stride 2, BN
5×5 conv. 32 ReLU, stride 2, BN	MLP output 3136, BN	5×5 conv. 64 ReLU, stride 2, BN
MLP output 784, BN	5×5 deconv. 64 ReLU, stride 2, BN	5×5 conv. 128 ReLU, stride 2, BN
		input z through MLP output 1024, ReLU
MLP output dim of z	5×5 deconv. 1 ReLU, stride 2, sigmoid	MLP output 1

Table 2: Architecture of the models for sVAE on CelebA. BN denotes batch normalization. lReLU denotes Leaky ReLU.

Encoder X to z	Decoder z to X	Discriminator
Input Image X concat with noise	Input z concat with noise	Input X
4×4 conv. 32 lReLU, stride 2, BN	concat random noise	5×5 conv. 64 ReLU, stride 2, BN
4×4 conv. 64 lReLU, stride 2, BN	MLP output 1024, lReLU, BN	5×5 conv. 128 ReLU, stride 2, BN
4×4 conv. 128 lReLU, stride 2, BN	MLP output 8192, lReLU, BN	5×5 conv. 256 ReLU, stride 2, BN
4×4 conv. 256 lReLU, stride 2, BN		5×5 conv. 512 ReLU, stride 2, BN
4×4 conv. 512 lReLU, stride 2, BN	5×5 deconv. 256 lReLU, stride 2, BN	Input z through MLP, output 2046, ReLU
MLP output 512, lReLU	5×5 deconv. 128 lReLU, stride 2, BN	concat two features from X and z
MLP output dim of z, tanh	5×5 deconv. 64 lReLU, stride 2, BN	
	5×5 deconv. 3 tanh, stride 2, BN	MLP output 1

Table 3: Architecture of the models for sVAE-r on CIFAR. BN denotes batch normalization. lReLU denotes Leaky ReLU. Dim denotes the number of attributes.

Encoder X to z	Decoder z to X	Discriminator
Input Image X concat with noise	Input z	Input X
5×5 conv. 32 lReLU, stride 2, BN	concat random noise	5×5 conv. 64 ReLU, stride 2, BN
5×5 conv. 64 lReLU, stride 2, BN		5×5 conv. 128 ReLU, stride 2, BN
5×5 conv. 128 lReLU, stride 2, BN	MP output 8192, lReLU, BN	5×5 conv. 256 ReLU, stride 2, BN
5×5 conv. 256 lReLU, stride 2, BN		5×5 conv. 512 ReLU, stride 2, BN, avg pooling
	5×5 deconv. 256 ReLU, stride 2, BN	Input z through MLP, output 512, ReLU
MLP output 512, lReLU	5×5 deconv. 128 ReLU, stride 2, BN	concat two features from X and z
MLP output dim of z, tanh	5×5 deconv. 3 tanh, stride 2	MLP output 1

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

2 More Result

2.1 CIFAR-10 result

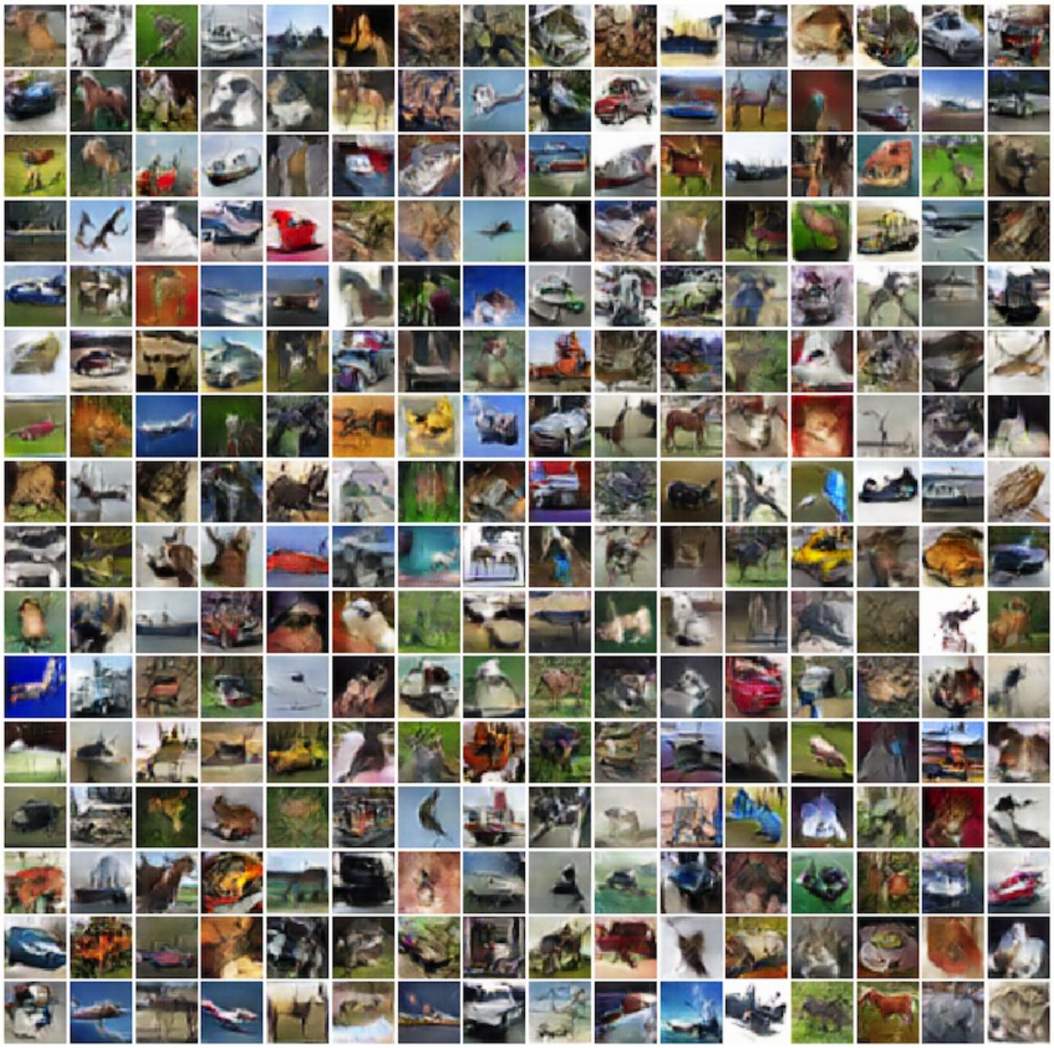

Figure 1: sVAE CIFAR unsupervised generation results with $\lambda = 0.1$.

2.2 CelebA result

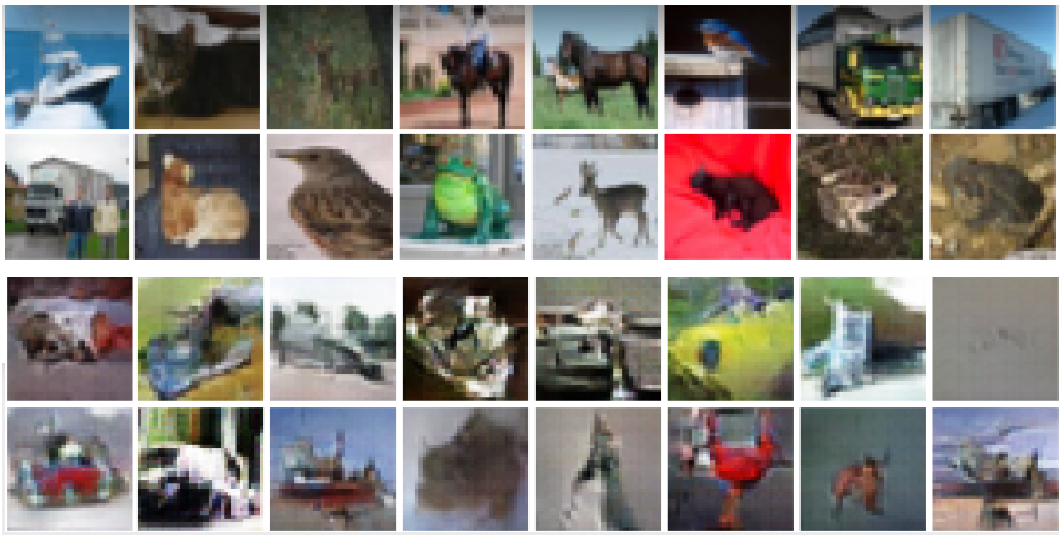

Figure 2: sVAE CIFAR unsupervised reconstruction. First two rows are original images, and the last two rows are the reconstructions

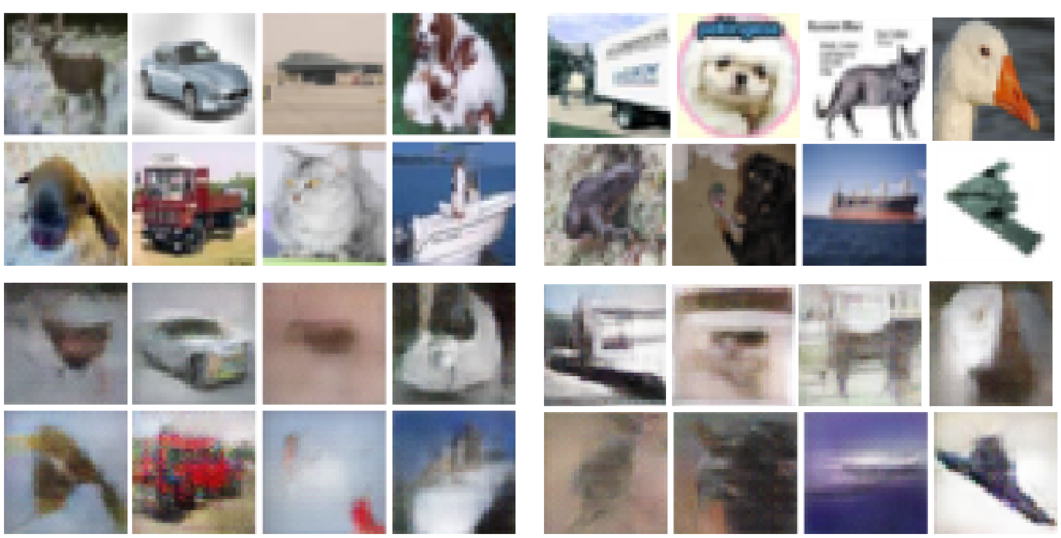

Figure 3: sVAE-r CIFAR unsupervised reconstruction. First two rows are original images, and the last two rows are the reconstructions

Figure 4: sVAE-r CelebA generations results with different λ

Figure 5: ALICE CelebA generations results with different λ

Figure 6: ALICE CelebA reconstructions with different λ.

Figure 7: ALICE CelebA reconstructions with different λ.