Symmetric Wasserstein Autoencoders (Supplementary File)

Sun Sun[1] **Hongyu Guo**[1]

[1]National Research Council Canada, Ottawa, ON., Canada

1 DATASETS AND NETWORK ARCHITECTURES

In this section, we briefly describe the datasets, the network architectures, and the hyperparameters that are used in our training algorithm.

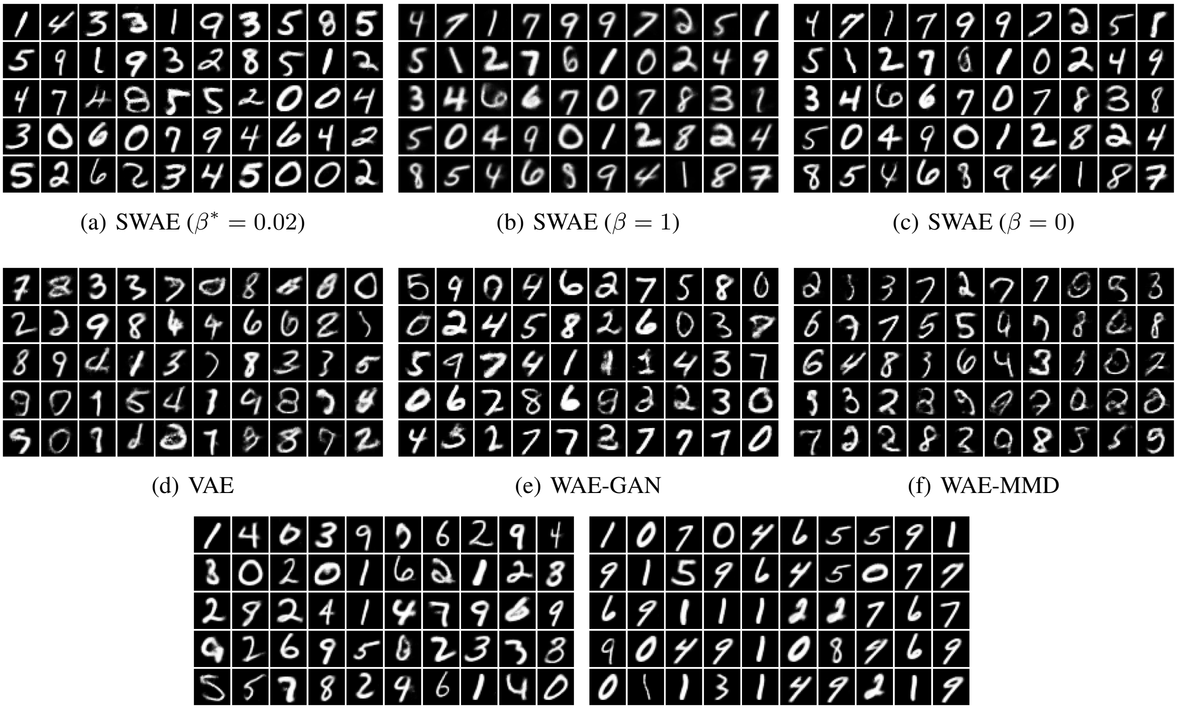
- MNIST: The dataset includes $70,000$ binarized images of numeric digits from 0 to 9, each of the size 28×28. There are $7,000$ images per class. The training set contains $50,000$ images, the validation set contains $10,000$ images for choosing the best model based on the loss function, and the test set contains $10,000$ images.
- Fashion-MNIST: The dataset includes $70,000$ binarized images of fashion products in 10 classes. This dataset has the same image size and the split of the training, validation, and test sets as in MNIST.
- Coil20: The dataset includes gray-scale images of 20 objects, each image of the size 32×32. The training set contains 1040 images, the validation set contains 200 images for choosing the best model based on the loss function, and the test set contains 200 images.
- CIFAR10-sub: The CIFAR-10 dataset consists of $60,000$ 32×32 colour images in 10 classes with $6,000$ images per class. There are $40,000$ training, $10,000$ validation, and $10,000$ test images. We randomly select three classes to form the CIFAR10-sub dataset, namely *bird, cat*, and *ship*.

Network architecture of SWAE: The building block of the network structure of SWAE is based on VampPrior, called GatedConv2d. GatedConv2d contains two convolutional layers with the gating mechanism utilized as an element-wise non-linearity. The parameters in the function GatedConv2d() represent the number of the input channels, the number of the output channels, kernel size, stride, and padding, respectively. The conditional prior network outputs the mean and the log-variance of a Gaussian distribution, based on which the latent prior is sampled.

- The structure of the encoder network: GatedConv2d(1,32,7,1,3)-GatedConv2d(32,32,3,2,1)-GatedConv2d(32,64,5,1,2)-GatedConv2d(64,64,3,2,1)-GatedConv2d(64,6,3,1,1), followed by one fully-connected layer with no activation function.
- The structure of the conditional prior network: The layers of GatedConv2d are the same as those in the encoder network, which are followed by two fully-connected layers. One produces the mean, and the other produces the log-variance with the activation function Hardtanh.
- The structure of the decoder network: Two fully-connected layers with the gating mechanism, followed by GatedConv2d(1,64,3,1,1)-GatedConv2d(64,64,3,1,1)-GatedConv2d(64,64,3,1,1)-GatedConv2d(64,64,3,1,1), followed by a convolutional layer with the activation function Sigmoid.

2 MORE EXPERIMENTAL RESULTS

In this section, we show more experimental results based on the comparison with the benchmarks.



(a) SWAE ($\beta^* = 0.02$)

(b) SWAE ($\beta = 1$)

(c) SWAE ($\beta = 0$)

(d) VAE

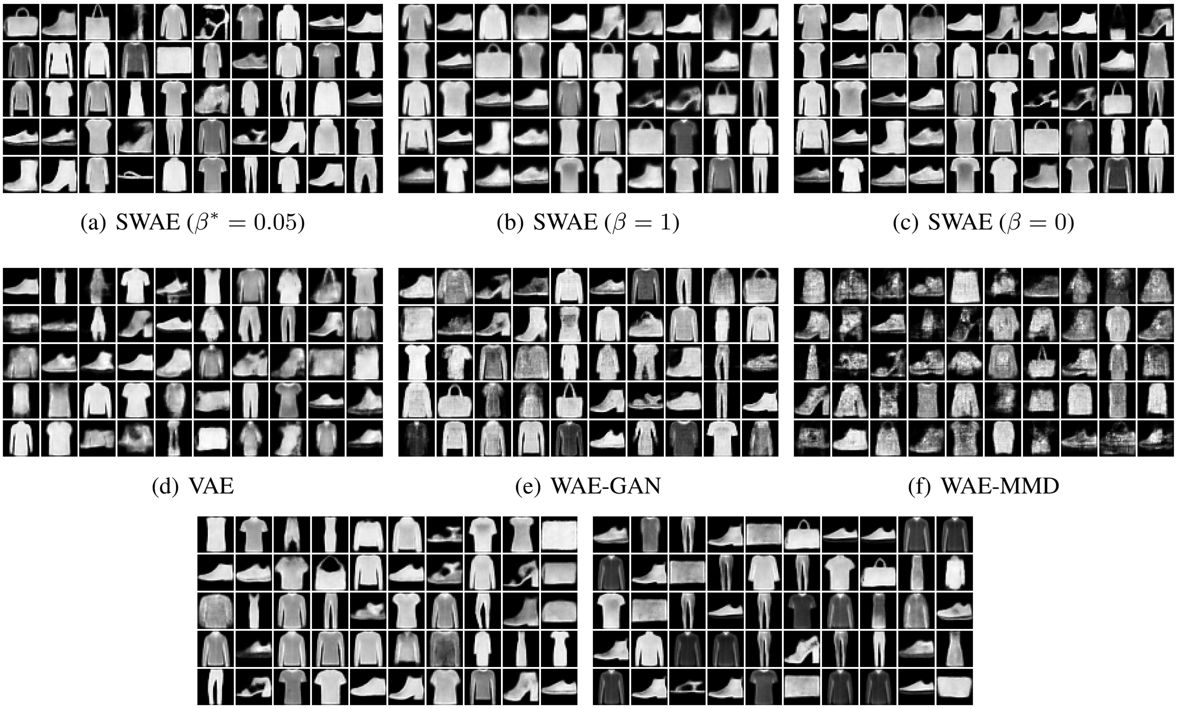
(e) WAE-GAN

(f) WAE-MMD

(g) VampPrior

(h) MIM

Figure 1: Generated new samples on MNIST. dim-$\mathbf{z} = 8$ for all methods.



(a) SWAE ($\beta^* = 0.05$)

(b) SWAE ($\beta = 1$)

(c) SWAE ($\beta = 0$)

(d) VAE

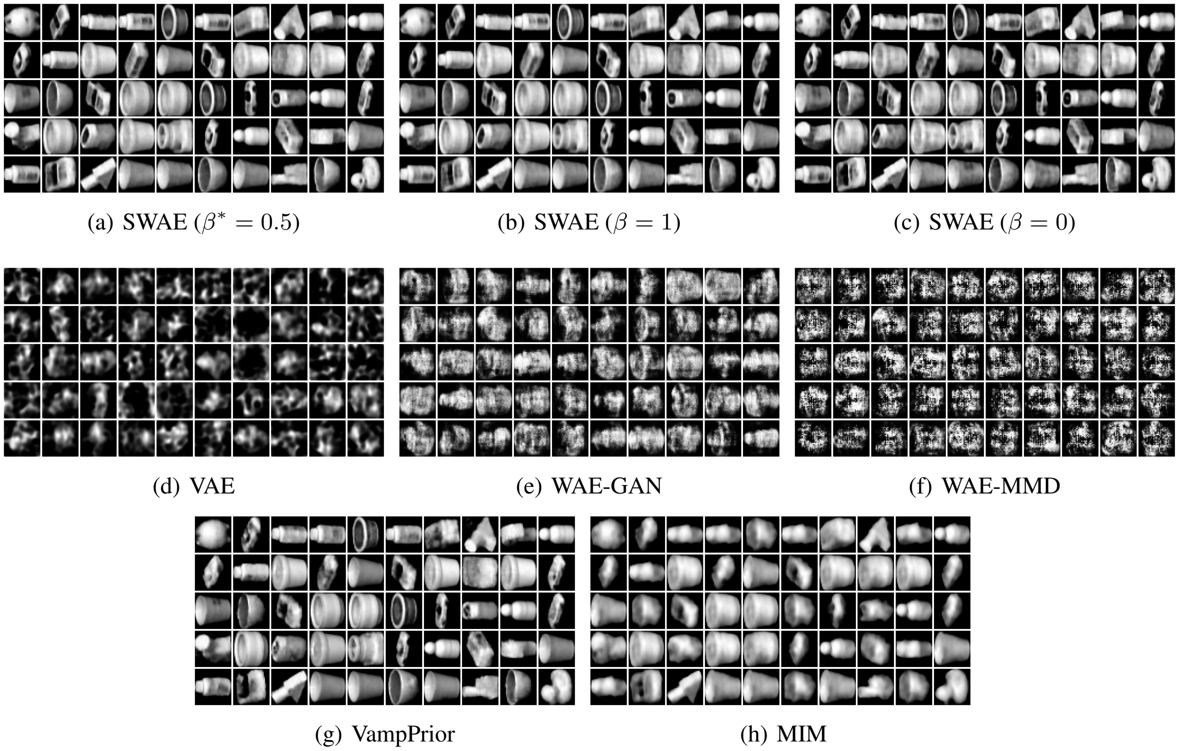
(e) WAE-GAN

(f) WAE-MMD

(g) VampPrior

(h) MIM

Figure 2: Generated new samples on Fashion-MNIST. dim-$\mathbf{z} = 8$ for all methods.



(a) SWAE ($\beta^* = 0.5$)

(b) SWAE ($\beta = 1$)

(c) SWAE ($\beta = 0$)

(d) VAE

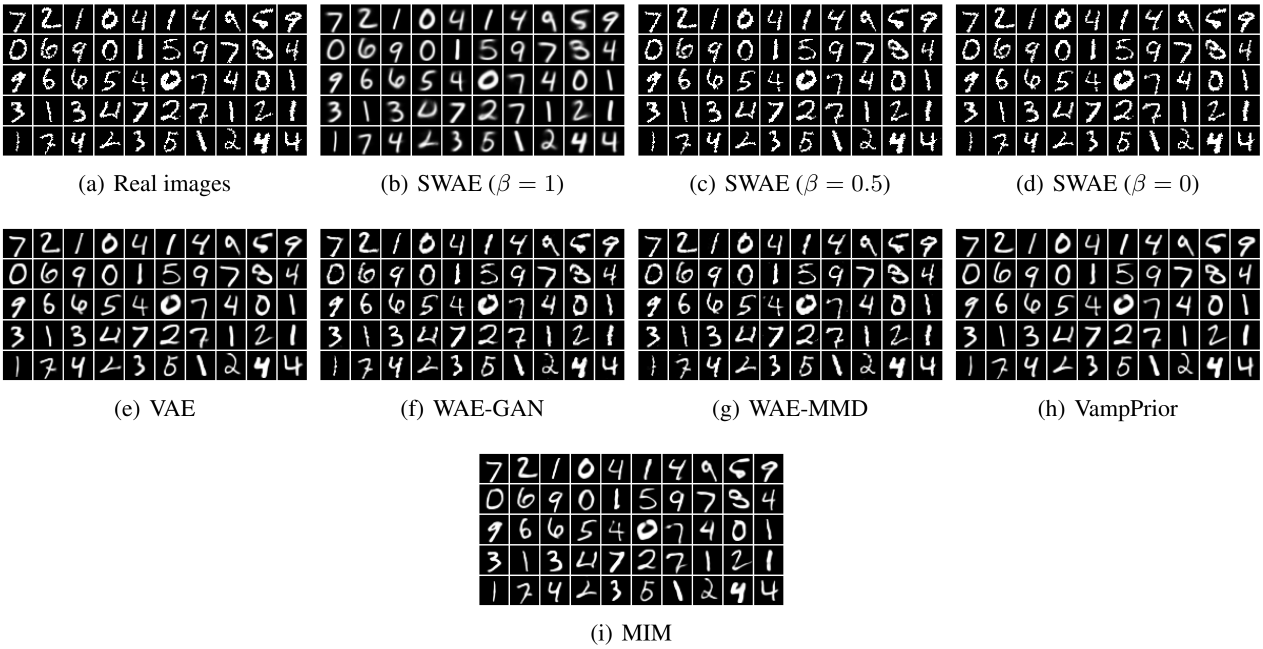
(e) WAE-GAN

(f) WAE-MMD

(g) VampPrior

(h) MIM

Figure 3: Generated new samples on Coil20. dim-$\mathbf{z} = 80$ for all methods.



(a) Real images

(b) SWAE ($\beta = 1$)

(c) SWAE ($\beta = 0.5$)

(d) SWAE ($\beta = 0$)

(e) VAE

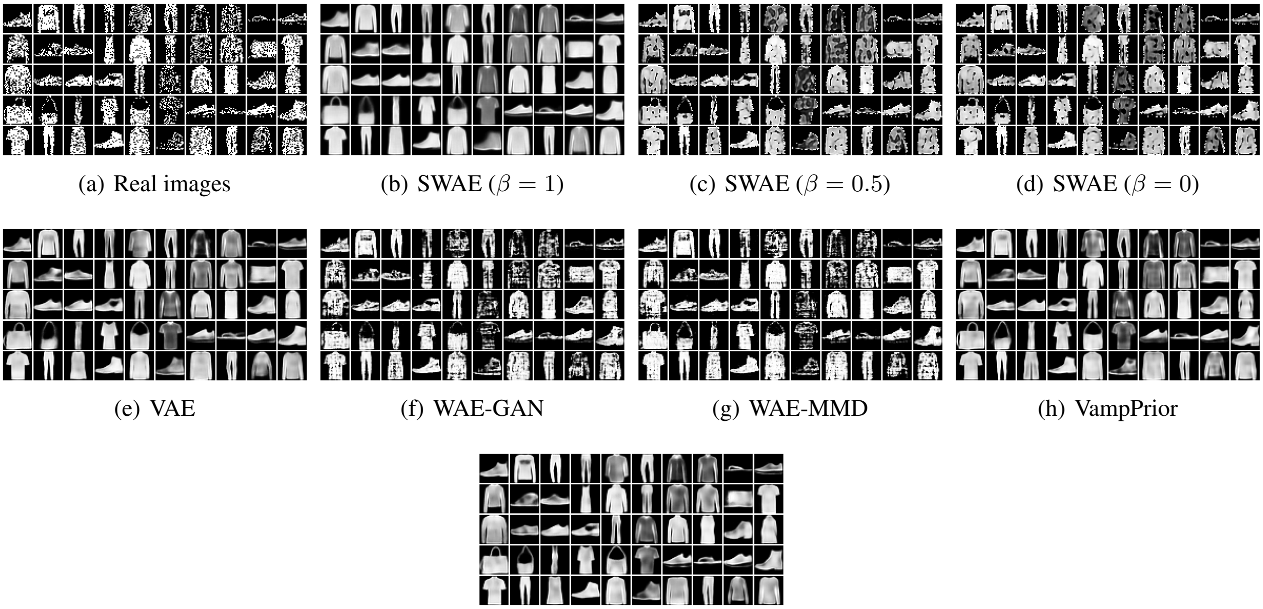
(f) WAE-GAN

(g) WAE-MMD

(h) VampPrior

(i) MIM

Figure 4: Reconstructed images on MNIST. dim-$\mathbf{z} = 80$ for all methods. As expected, for SWAEs a smaller β leads to a higher quality of reconstruction.



(a) Real images

(b) SWAE ($\beta = 1$)

(c) SWAE ($\beta = 0.5$)

(d) SWAE ($\beta = 0$)

(e) VAE

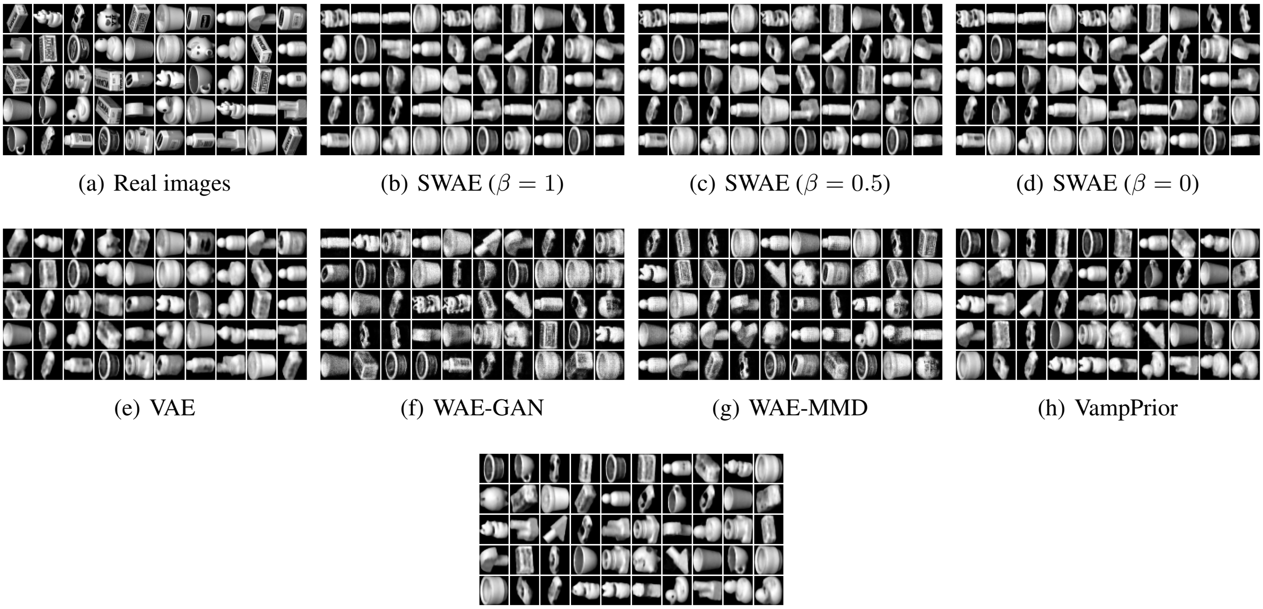
(f) WAE-GAN

(g) WAE-MMD

(h) VampPrior

(i) MIM

Figure 5: Reconstructed images on Fashion-MNIST. dim-$\mathbf{z} = 80$ for all methods.



(a) Real images

(b) SWAE ($\beta = 1$)

(c) SWAE ($\beta = 0.5$)

(d) SWAE ($\beta = 0$)

(e) VAE

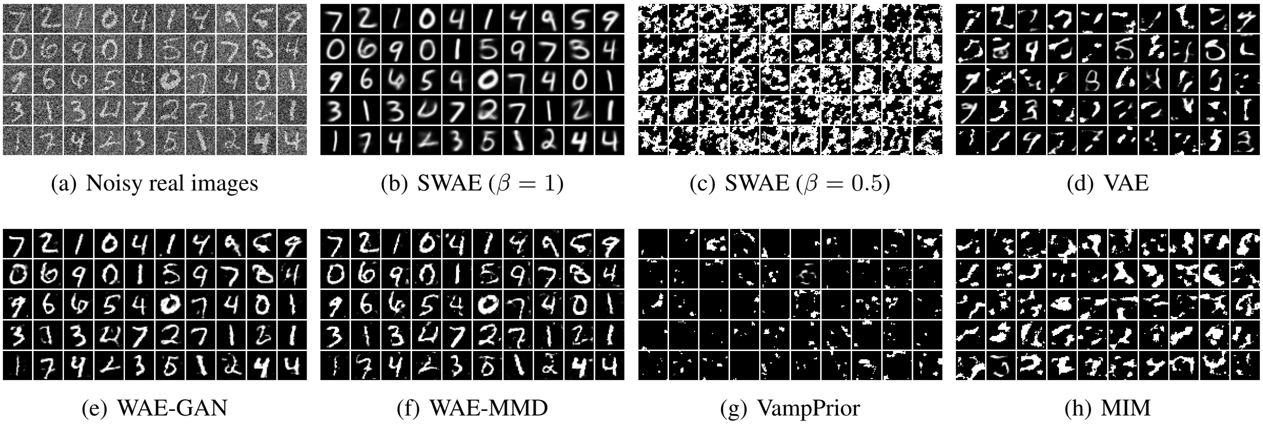
(f) WAE-GAN

(g) WAE-MMD

(h) VampPrior

(i) MIM

Figure 6: Reconstructed images on Coil20. dim-$\mathbf{z} = 80$ for all methods. For SWAEs, the difference of the reconstruction error for different values of β is insignificant, and the reconstructed images look visually the same.



(a) Noisy real images

(b) SWAE ($\beta = 1$)

(c) SWAE ($\beta = 0.5$)

(d) VAE

(e) WAE-GAN

(f) WAE-MMD

(g) VampPrior

(h) MIM

Figure 7: Denoising effect: reconstructed images on MNIST. dim-$\mathbf{z} = 80$ for all methods. SWAE ($\beta = 1$), WAE-GAN, and WAE-MMD can recover clean images. However, for WAE-GAN and WAE-MMD, we can still see some noisy dots around the digits.