Do Bayesian Variational Autoencoders Know What They Don't Know? (Supplementary material)

Misha Glazunov[1] **Apostolis Zarras**[1]

[1]Delft University of Technology, the Netherlands

A SAMPLE STANDARD DEVIATIONS OF THE MARGINAL LOG-LIKELIHOODS

The sample standard deviations of the marginal log-likelihoods for BBB and SGHMC methods can be observed in Figure 1.

B VAE DISTRIBUTIONS

- For prior we used a standard multivariate Gaussian without parameters: $p(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{0}, \mathbf{I})$
- For variational distribution we used a multivariate factorized Gaussian with learned mean and variance: $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}; \boldsymbol{\mu}, diag(\boldsymbol{\sigma}^{\mathbf{2}}))$
- For likelihood we used a multivariate factorized Bernoulli distribution:

$$p(\mathbf{x} \mid \mathbf{z}) = \prod_{j=1}^{D} p\left(x_j \mid \mathbf{z}\right) = \prod_{j=1}^{D} \operatorname{Bernoulli}\left(x_j; p_j\right) \tag{1}$$

C CNN ARCHITECTURES USED

For MNIST and FashionMNIST datasets with a single channel we used the following architectures depicted in Table 1 and in Table 2.

Table 1: Encoder CNN for MNIST and FashionMNIST

Operation	**Kernel**	**Strides**	**Feature Maps**
Convolution	3 x 3	1 x 1	32
Convolution	3 x 3	1 x 1	16
Max pooling 2D	2 x 2	2 x 2	—
Linear for $\boldsymbol{\mu}$	—	—	10
Linear for $\log \boldsymbol{\sigma}$	—	—	10

Table 2: Decoder CNN for MNIST and FashionMNIST

Operation	**Kernel**	**Strides**	**Feature Maps**
Linear for sampled $\mathbf{z}$	—	—	2306
Upsampling nearest 2D	—	—	—
Max pooling 2D	2 x 2	2 x 2	—
Transposed Convolution	3 x 3	1 x 1	32
Transposed Convolution	3 x 3	1 x 1	1

For SVHN and CIFAR10 datasets with three channels we used the following architectures with additional padding = 1 and no bias for every convolutional layer (see Table 3 and Table 4). For SVHN latent dimensionality = 20, for CIFAR10 = 70.

Table 3: Encoder CNN for SVHN and CIFAR10

Operation	**Kernel**	**Strides**	**Feature Maps**
Convolution	3 x 3	1 x 1	16
Batch normalization	—	—	16
Convolution	3 x 3	2 x 2	32
Batch normalization	—	—	32
Convolution	3 x 3	1 x 1	32
Batch normalization	—	—	32
Convolution	3 x 3	2 x 2	16
Batch normalization	—	—	16
Linear	—	—	512
Batch normalization	—	—	512
Linear for $\boldsymbol{\mu}$	—	—	20 / 70
Linear for $\log \boldsymbol{\sigma}$	—	—	20 / 70

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

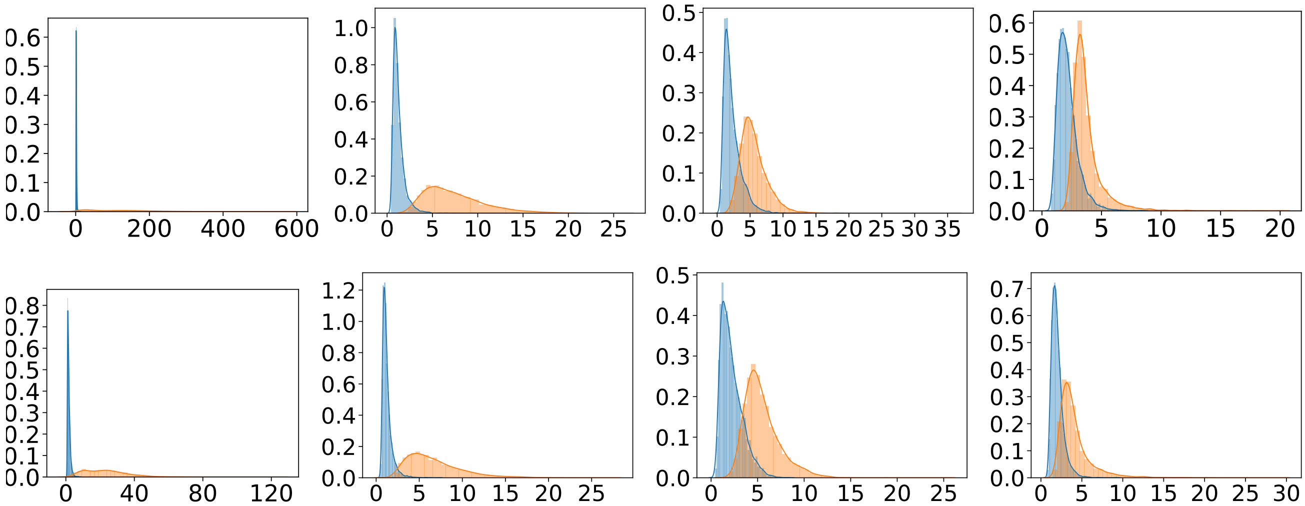

Figure 1: Histograms of the sample standard deviations of the marginal log-likelihoods, blue depicts in-distribution (ID) and orange - out-of-distribution (OoD). **From left to right**: MNIST as ID vs Fashion-MNIST as OoD, Fashion-MNIST as ID vs MNIST as OoD, SVHN as ID vs CIFAR-10 as Ood, CIFAR-10 as ID vs SVHN as OoD. **Top:** Sampling is done from Bayes-by-backprop VAE. **Bottom:** Sampling is done from SGHMC VAE.

Table 4: Decoder CNN for SVNH and CIFAR10

Operation	**Kernel**	**Strides**	**Feature Maps**
Linear for sampled $\mathbf{z}$	—	—	512
Batch normalization	—	—	512
Linear	—	—	1024
Batch normalization	—	—	1024
Transposed Convolution	3 x 3	2 x 2	32
Batch normalization	—	—	32
Transposed Convolution	3 x 3	1 x 1	32
Batch normalization	—	—	32
Transposed Convolution	3 x 3	2 x 2	16
Batch normalization	—	—	16
Transposed Convolution	3 x 3	1 x 1	3

For all architectures we used ReLU as a non-linearity. In addition, all pixels of the images have been normalized to [0,1] range for each channel for both training and testing phases.

D SAMPLES FROM TRAINED MODELS

Random samples from all of the trained models for both BBB and SGHMC can be seen on Figure 2.

E RUNTIMES OF DIFFERENT METHODS

The runtimes for the training convergence for CIFAR-10 (the most complex dataset used in the experiments) for different *Bayesian* methods are available in Table 5

Table 5: BVAE runtimes for learning

Method	**Time (mins)**
BBB	1628
SGHMC	1473
SWAG	371
Vanilla	345

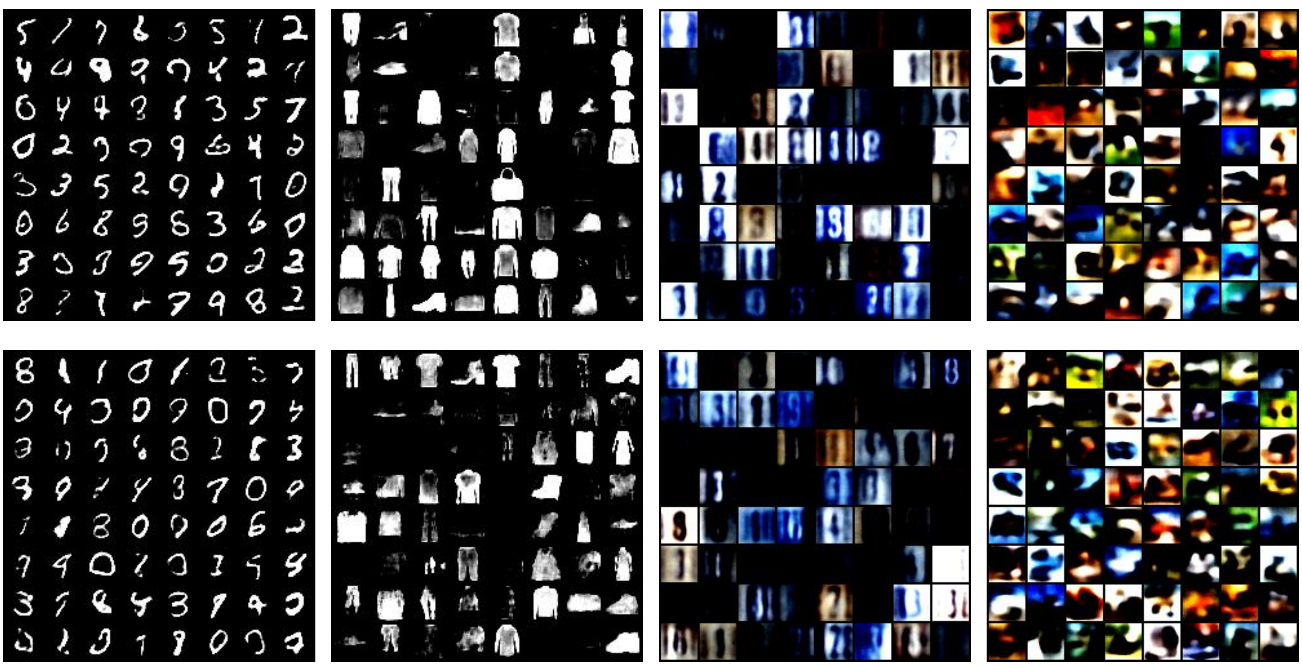

Figure 2: **From left to right**: MNIST, Fashion-MNIST, SVHN, CIFAR-10. **Top:** Random samples from BBB VAE. **Bottom:** Random samples from SGHMC VAE.