# Towards Bridging the Performance Gaps of Joint Energy-based Models Supplementary Material

Xiulong Yang, Qing Su, and Shihao Ji
Georgia State University
{xyang22,qsu3,sji}@gsu.edu

## A. Experimental Details

To have a fair comparison, we largely follow the settings of JEM [2] and JEM++ [5], and train our models based on the Wide-ResNet 28x10 architecture [6] for 200 epochs. We use SGD for CIFAR10 and CIFAR100 with an initial learning rate of 0.1 and 0.01, respectively, and decay the learning rate by 0.2 at epoch [60, 120, 180] for most cases. Apart from this, we find that the cosine learning rate scheduler can be adopted for SADA-JEM, which achieves much better accuracy and FID on CIFAR10.[1] The hyper-parameters used in our experiments are listed in Table 1.

Table 1. Hyper-parameters of SADA-JEM for CIFAR10 and CIFAR100.

| Variable | Value |
|---|---|
| Number of SGLD steps $K$ | 5, 10, 20 |
| Buffer size $\lvert\mathbb{B}\rvert$ | 10,000 |
| Reinitialization freq. $\gamma$ | 5% |
| SGLD step-size $\alpha$ | 1 |
| SGLD noise $\sigma$ | 0 |
| SAM noise radius $\rho$ | 0.2 |

## B. Visualizing Generated Images

Table 1 in the main text reports the quantitative performance comparison of different stand-alone generative models and hybrid models. Here in Figure 1 we provide a qualitative comparison of generated images from (a) SADA-JEM, (b) VERA [3], and (c) DiffuRecov [1]. As we can see, the perceived image qualities of them are comparable even though DiffuRecov has a much better FID score than that of VERA (9.58 vs. 30.5), indicating that visualizing generated images is less effective to evaluate image quality.

[1] This is because the combination of SAM and single branched DA improves the training stability significantly. As a result, the cosine learning rate decay can be adopted to improve the overall performance. JEM, JEM++ and other SADA-JEM ablation configurations are less stable to enable the cosine learning rate decay.

## C. Energy Landscapes

Figure 2 illustrates the energy landscapes of different models trained on CIFAR10. The energy landscape is generated by visualizing $E(\boldsymbol{\theta}) = \sum_{\boldsymbol{x}\in \boldsymbol{X}} E_{\boldsymbol{\theta}}(\boldsymbol{x})$ with the technique introduced in [4], where $X$ is a 10% random samples from CIFAR10 training data. As we can see, SADA-JEM's energy landscapes are much smoother than those of the competing methods (see different scales of the y-axes).

## D. Out-of-Distribution Detection

Table 2 reports the OOD detection performances of different models and SADA-JEM with different $K$s, where the input density $\log p_{\boldsymbol{\theta}}(\boldsymbol{x})$ is used as $s_{\boldsymbol{\theta}}(\boldsymbol{x})$ for OOD detection on CIFAR10.

## E. Additional Generated Samples

Additional SADA-JEM generated class-conditional (best and worst) samples of CIFAR10 are provided in Figures 3-12.

## References

[1] Ruiqi Gao, Yang Song, Ben Poole, Ying Nian Wu, and Diederik P. Kingma. Learning Energy-Based Models by Diffusion Recovery Likelihood. In *ICLR*, 2021. 1

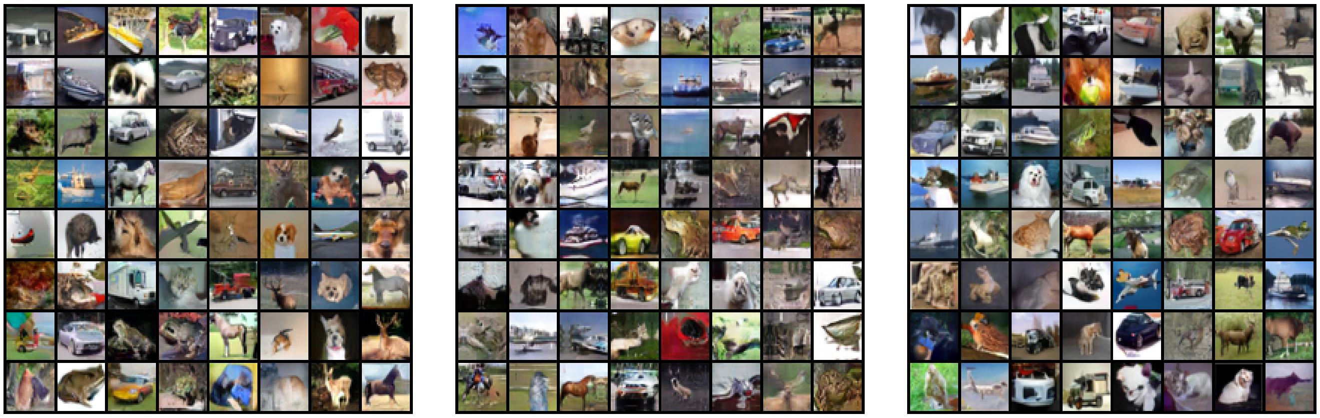
[2] Will Grathwohl, Kuan-Chieh Wang, Joern-Henrik Jacobsen, David Duvenaud, Mohammad Norouzi, and Kevin Swersky. Your classifier is secretly an energy based model and you should treat it like one. In *International Conference on Learning Representations (ICLR)*, 2020. 1

[3] Will Sussman Grathwohl, Jacob Jin Kelly, Milad Hashemi, Mohammad Norouzi, Kevin Swersky, and David Duvenaud. No mcmc for me: Amortized sampling for fast and stable training of energy-based models. In *ICLR*, 2021. 1

[4] Hao Li, Zheng Xu, Gavin Taylor, Christoph Studer, and Tom Goldstein. Visualizing the Loss Landscape of Neural Nets. In *Neural Information Processing Systems (NeurIPS)*, 2018. 1

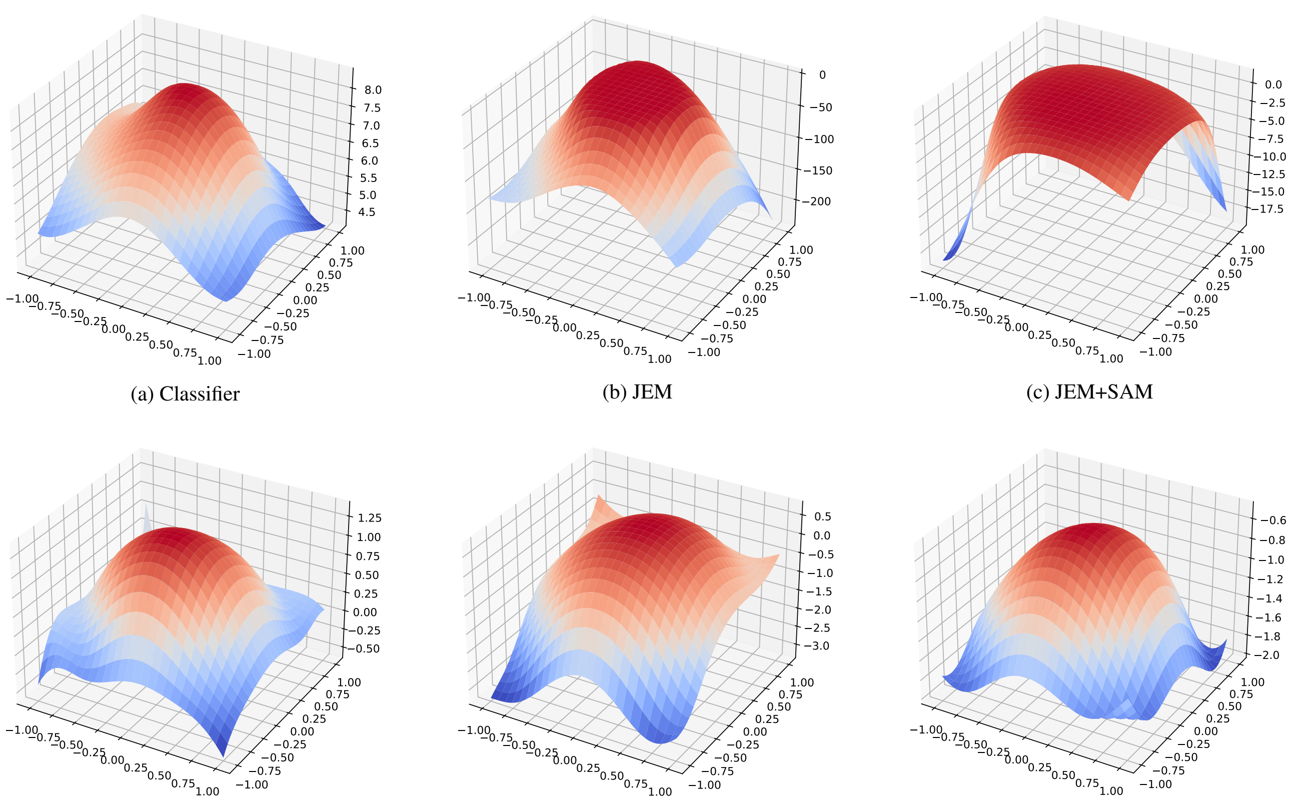
[5] Xiulong Yang and Shihao Ji. JEM++: Improved Techniques for Training JEM. In *International Conference on Computer Vision (ICCV)*, 2021. 1

[6] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *BMVC*, 2016. 1



(a) SADA-JEM (b) VERA (c) DiffuRecov

Figure 1. Generated images from SADA-JEM, VERA, and DiffuRecov.



(a) Classifier (b) JEM (c) JEM+SAM

(d) SADA-JEM (K=5) (e) SADA-JEM (K=10) (f) SADA-JEM (K=20)

Figure 2. Energy landscapes of different models trained on CIFAR10. Please note the different scales of the y-axes.

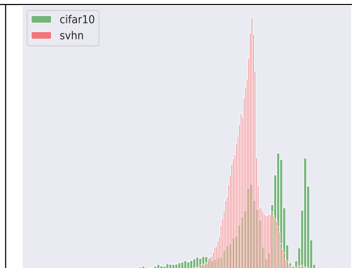
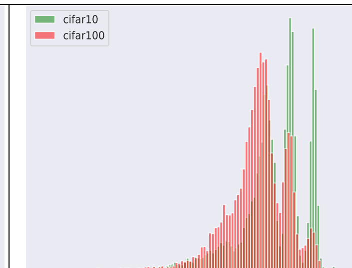
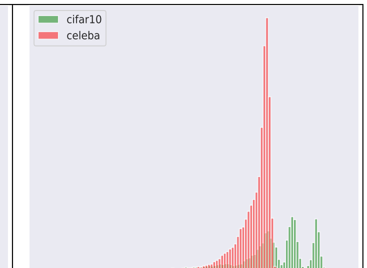
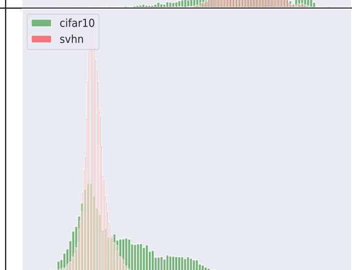
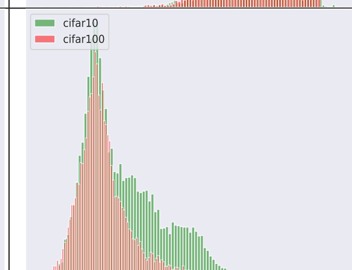
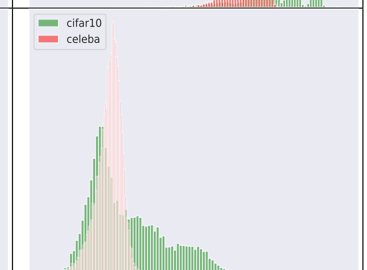
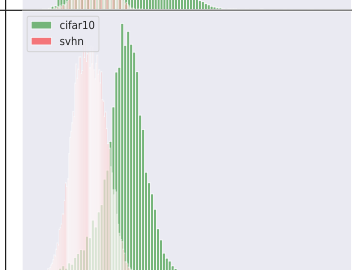
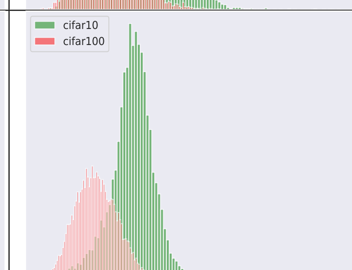
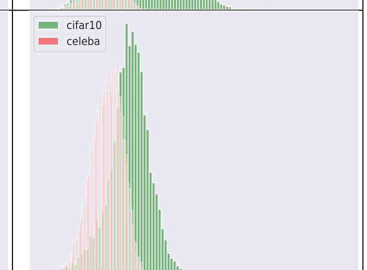
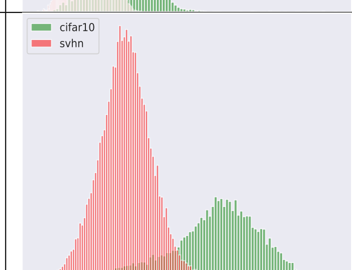
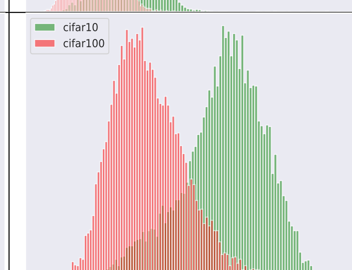
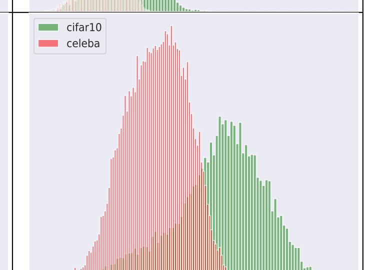
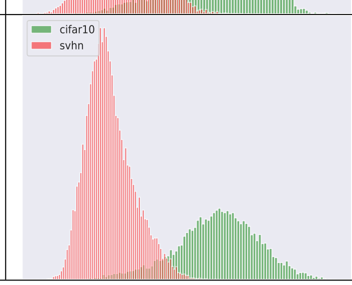
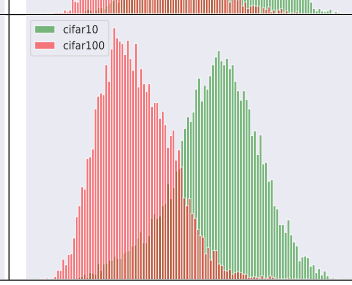
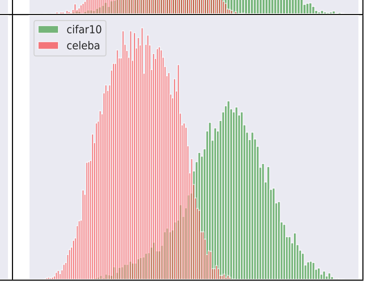
| | | | |
|---|---|---|---|
| JEM |  |  |  |
| JEM++ (M=10) |  |  |  |
| SADA-JEM (K=5) |  |  |  |
| SADA-JEM (K=10) |  |  |  |
| SADA-JEM (K=20) |  |  |  |

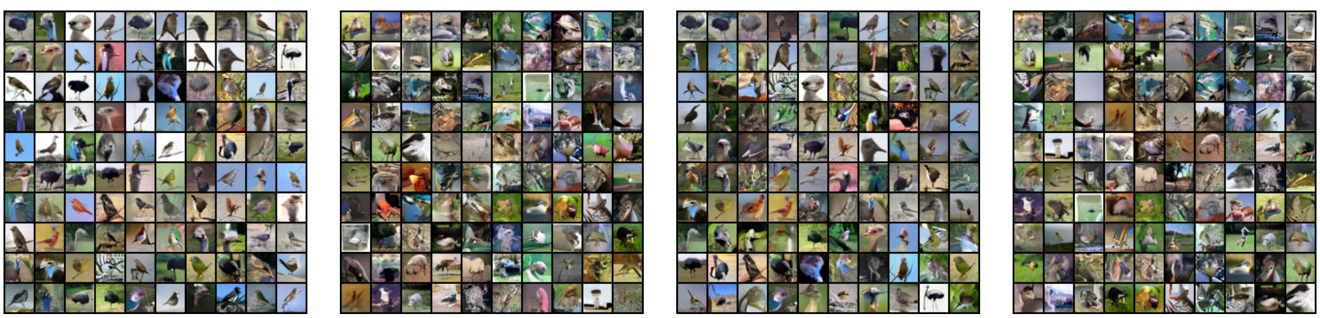
Table 2. Histograms of $\log p_{\boldsymbol{\theta}}(\boldsymbol{x})$ for OOD detection. Green corresponds to in-distribution dataset, while red corresponds to OOD dataset.

(a) Samples with highest $p(\boldsymbol{x})$ (b) Samples with lowest $p(\boldsymbol{x})$ (c) Samples with highest $p(y|\boldsymbol{x})$ (d) Samples with lowest $p(y|\boldsymbol{x})$

Figure 3. SADA-JEM generated class-conditional samples of **Plane**.

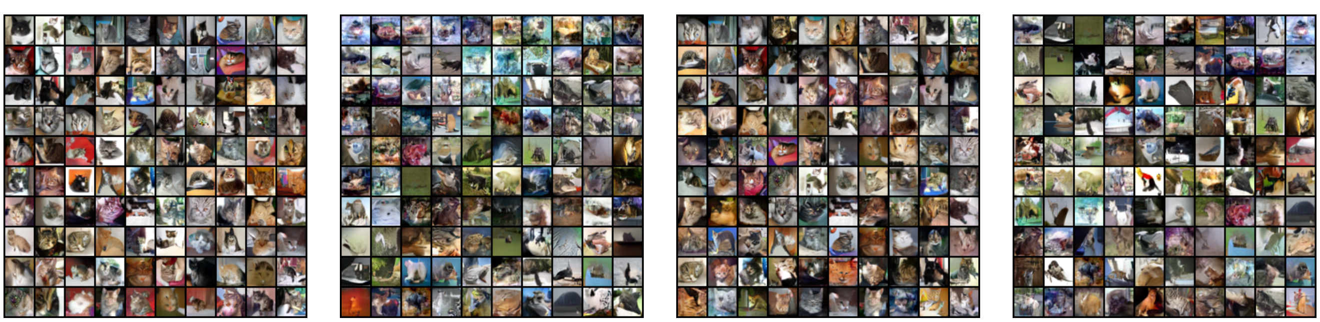
(a) Samples with highest $p(\boldsymbol{x})$ (b) Samples with lowest $p(\boldsymbol{x})$ (c) Samples with highest $p(y|\boldsymbol{x})$ (d) Samples with lowest $p(y|\boldsymbol{x})$

Figure 4. SADA-JEM generated class-conditional samples of **Car**.



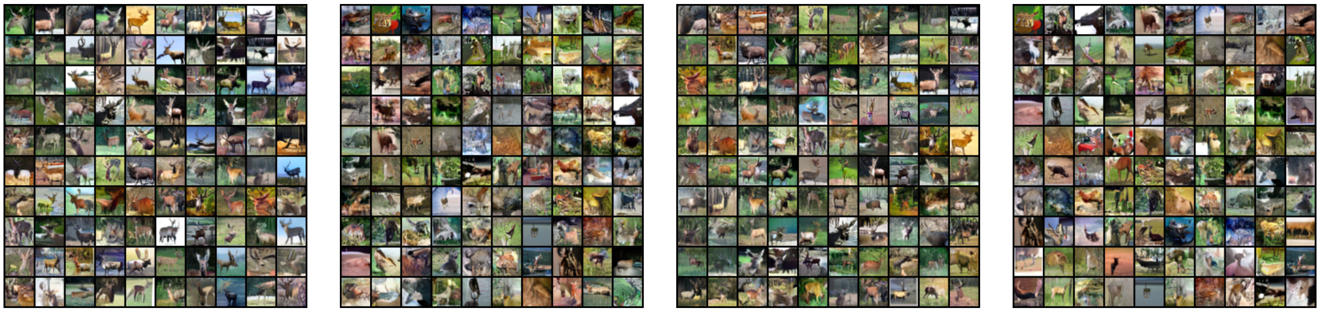
(a) Samples with highest $p(\boldsymbol{x})$ (b) Samples with lowest $p(\boldsymbol{x})$ (c) Samples with highest $p(y|\boldsymbol{x})$ (d) Samples with lowest $p(y|\boldsymbol{x})$

Figure 5. SADA-JEM generated class-conditional samples of **Bird**.



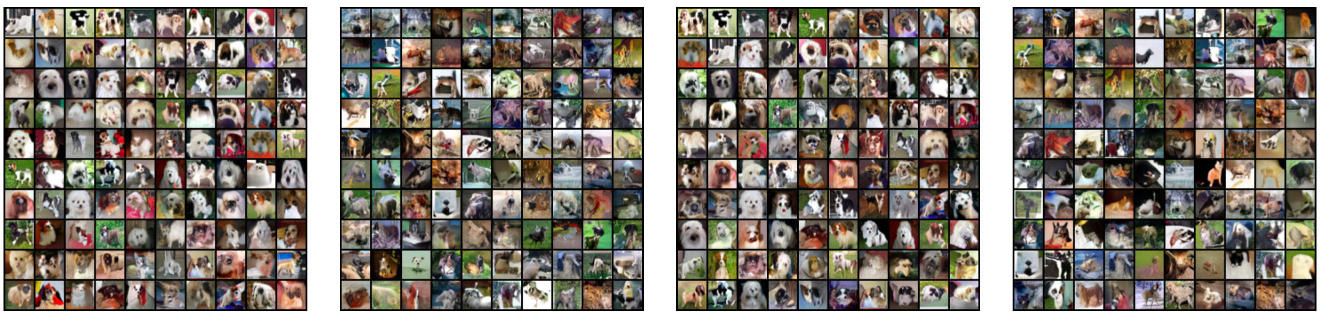
(a) Samples with highest $p(\boldsymbol{x})$ (b) Samples with lowest $p(\boldsymbol{x})$ (c) Samples with highest $p(y|\boldsymbol{x})$ (d) Samples with lowest $p(y|\boldsymbol{x})$

Figure 6. SADA-JEM generated class-conditional samples of **Cat**.



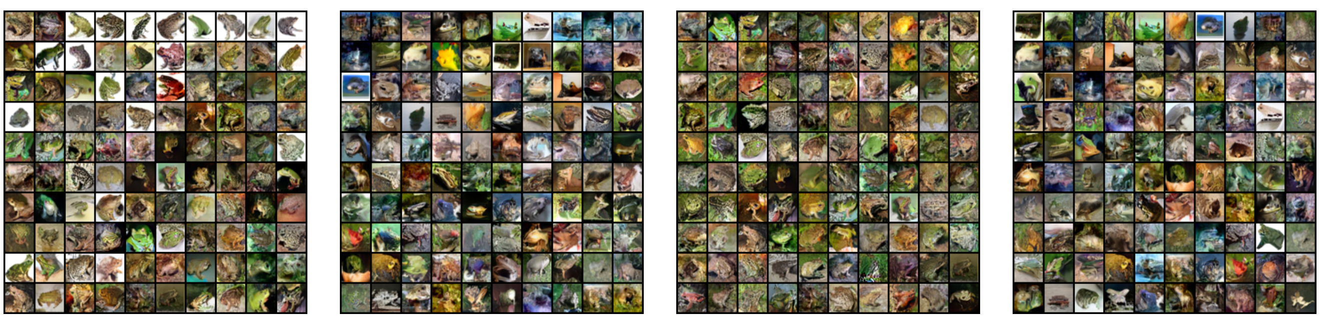
(a) Samples with highest $p(\boldsymbol{x})$ (b) Samples with lowest $p(\boldsymbol{x})$ (c) Samples with highest $p(y|\boldsymbol{x})$ (d) Samples with lowest $p(y|\boldsymbol{x})$

Figure 7. SADA-JEM generated class-conditional samples of **Deer**.



(a) Samples with highest $p(\boldsymbol{x})$ (b) Samples with lowest $p(\boldsymbol{x})$ (c) Samples with highest $p(y|\boldsymbol{x})$ (d) Samples with lowest $p(y|\boldsymbol{x})$

Figure 8. SADA-JEM generated class-conditional samples of **Dog**.

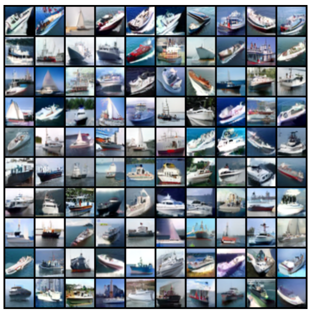


(a) Samples with highest $p(\boldsymbol{x})$ (b) Samples with lowest $p(\boldsymbol{x})$ (c) Samples with highest $p(y|\boldsymbol{x})$ (d) Samples with lowest $p(y|\boldsymbol{x})$

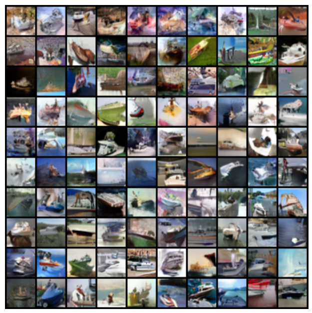
Figure 9. SADA-JEM generated class-conditional samples of **Frog**.

(a) Samples with highest $p(\boldsymbol{x})$ (b) Samples with lowest $p(\boldsymbol{x})$ (c) Samples with highest $p(y|\boldsymbol{x})$ (d) Samples with lowest $p(y|\boldsymbol{x})$

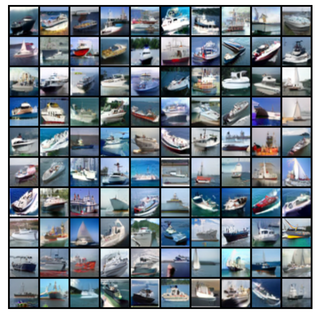
Figure 10. SADA-JEM generated class-conditional samples of **Horse**.



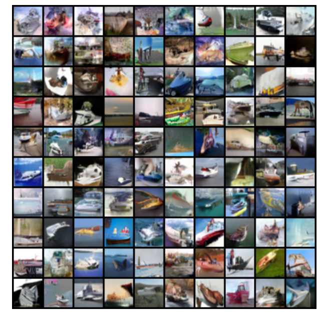
(a) Samples with highest $p(\boldsymbol{x})$



(b) Samples with lowest $p(\boldsymbol{x})$

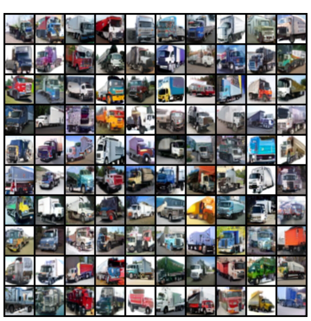


(c) Samples with highest $p(y|\boldsymbol{x})$

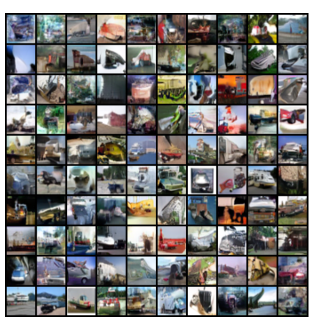


(d) Samples with lowest $p(y|\boldsymbol{x})$

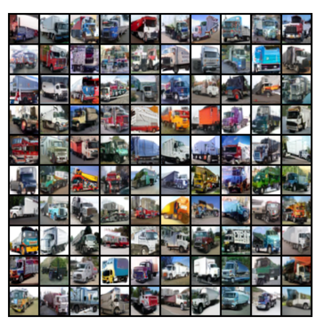
Figure 11. SADA-JEM generated class-conditional samples of **Ship**.



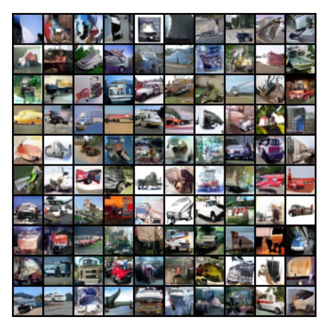
(a) Samples with highest $p(\boldsymbol{x})$



(b) Samples with lowest $p(\boldsymbol{x})$



(c) Samples with highest $p(y|\boldsymbol{x})$



(d) Samples with lowest $p(y|\boldsymbol{x})$

Figure 12. SADA-JEM generated class-conditional samples of **Truck**.