Deep Compressed Sensing: Appendix

Yan Wu [1] **Mihaela Rosca** [1] **Timothy Lillicrap** [1]

1. Experiments Detail

Unless otherwise specified, we used the following default configuration for all experiments. We used the Adam optimiser (Kingma & Ba, 2014) with the learning rate 1×10^{-4} and the parameters $\beta_1 = 0.9$, $\beta_2 = 0.999$. We trained all the models for 4×10^5 steps with the batch size of 64. 100 dimensional latent representations were used for generators. We use 3 gradient-descent steps for latent optimisation, and the initial step size of 0.01.

1.1. Reconstruction Experiments

Following Bora et al. (2017), we used a 2-layer multi-layer perceptron (MLP), with 500 units in each hidden layer and leaky ReLU non-linearity, as the generator for MNIST images; for CelebA, we used the DCGAN generator (Radford et al., 2015). In addition to random linear projections, we tested the following neural networks as the measurement functions: a 2-layer MLP with 500 units in each layer and leaky ReLU non-linearity for MNIST, and the DCGAN discriminator for CelebA.

1.2. GAN experiments

We used the same MLP generator and discriminator (i.e., measurement function) as described in the previous section for MNIST experiments. We also use the same architecture for the semi-supervised GAN.

For CIFAR dataset, we used the DCGAN architecture with its recommended Adam parameters $\beta_1 = 0.5$, $\beta_2 = 0.9$ (Radford et al., 2015). We tested a number of hyper-parameters as the cross product of the following: generator learning rates $\{1 \times 10^{-4}, 2 \times 10^{-4}, 3 \times 10^{-4}\}$, discriminator learning rates $\{1 \times 10^{-4}, 2 \times 10^{-4}, 3 \times 10^{-4}\}$, latent variable sizes $\{100, 200\}$, mini-batch sizes $\{32, 64\}$. Additionally, 2 replicas for each combination were trained to account for the effect of random seeds.

To reproduce the Spectral Normalised GANs. We used the same discriminator as in Miyato et al. (2018), which is deeper than the DCGAN discriminator. A grid search over optimisation parameters found the learning rate of 1×10^{-4} and Adam's β_2 of 0.999 most stably achieved the best results.

Inception Scores and Frchet Inception Distances were reported as the averages of 10 evaluations each based on $5,000$ random samples (Salimans et al., 2016; Heusel et al., 2017).

[1]DeepMind, London, UK. Correspondence to: Yan Wu <yanwu@google.com>.

Proceedings of the 36th International Conference on Machine Learning, Long Beach, California, PMLR 97, 2019. Copyright 2019 by the author(s).

References

Bora, A., Jalal, A., Price, E., and Dimakis, A. G. Compressed sensing using generative models. *arXiv preprint arXiv:1703.03208*, 2017.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., and Hochreiter, S. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pp. 6626–6637, 2017.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Miyato, T., Kataoka, T., Koyama, M., and Yoshida, Y. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations*, 2018. URL `https://openreview.net/forum?id=B1QRgziT-`.

Radford, A., Metz, L., and Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., and Chen, X. Improved techniques for training gans. In *Advances in Neural Information Processing Systems*, pp. 2234–2242, 2016.