Supplemental Materials: TREND: Truncated Generalized Normal Density Estimation of Inception Embeddings for GAN Evaluation

Junghyuk Lee and Jong-Seok Lee

School of Integrated Technology, Yonsei University, Seoul, Republic of Korea
{junghyuklee,jong-seok.lee}@yonsei.ac.kr

A Inter-dimensional independence of Inception features

Fig. A shows the histogram of the PCC value between each pair of dimensions of Inception features, which shows near independence between feature dimensions (Section 3). The scatter plots in Fig. B also confirm approximate independence between dimensions.

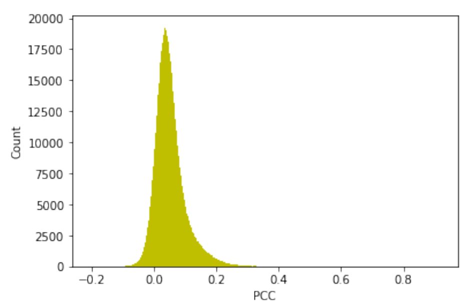


Fig. A: Histogram of PCCs between pairs of Inception feature dimensions.

B Experimental Setup

For target images, we use the CIFAR10 [8], CelebA [9], Flicker-Face-HQ (FFHQ) [6], and ImageNet [2] datasets. We use the validation split of the datasets or the test split only if the validation split is not provided. For evaluation of generative models, we generate images using the pre-trained models as follows: DCGAN [10] trained on the CIFAR10 dataset, ProGAN [5] trained on the CelebA dataset, BigGAN-deep-256 [1], ADM [3], E-VDVAE [4], trained on the ImageNet dataset, and StyleGAN [6], StyleGAN2 [7], and E-VDVAE [4] trained on the FFHQ dataset.

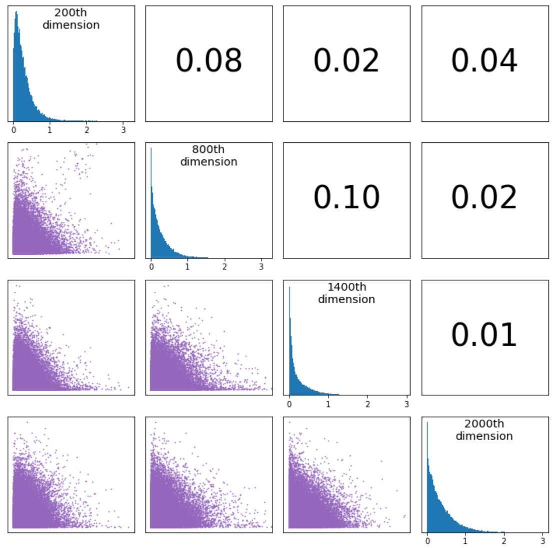
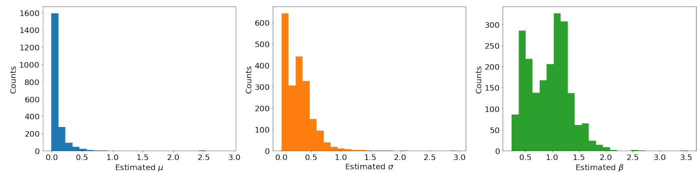


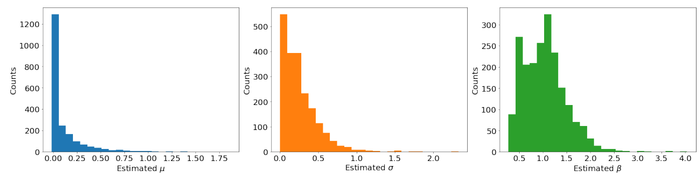
Fig. B: Scatter plots showing the inter-dimensional relationships of Inception features for selected feature dimensions (200th, 800th, 1400th, and 2000th). The corresponding PCC values are also shown. The histogram of each dimension is shown on the diagonal.

We use the pre-trained DCGAN [10] model[1] trained on CIFAR10 to generate 32×32 images. We generate 256×256 images using the pre-trained BigGAN [1] model[2] trained on ImageNet. Since BigGAN is trained on the conditional class label of ImageNet, we generate a fixed number (50) of images for each class. We also apply the truncation trick for BigGAN with varying the threshold value of latent vectors. In general, a smaller threshold value yields a lower level of diversity of the generated images. We use five different threshold values: 0.2, 0.4, 0.6, 0.8, and ∞ (i.e., no truncation trick). We use StyleGAN [6] that is trained on FFHQ to generate high-resolution face images having a resolution of 1024×1024 pixels. We use samples downloaded from its official website[3]. For StyleGAN2, We use the pre-trained model[4] [7] that is also trained on the FFHQ dataset. We use the pre-trained ADM-C and ADM-U [10] models[5] trained on ImageNet with classifier guidance and up-sampling, respectively. For E-VDVAE [4], we use the pre-trained models[6] trained on ImageNet and FFHQ datasets.

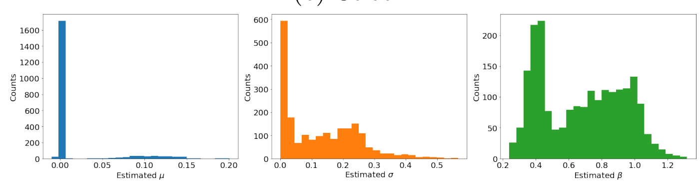
[1] https://github.com/csinva/gan-vae-pretrained-pytorch
[2] https://tfhub.dev/deepmind/biggan-deep-256/1
[3] https://github.com/NVlabs/stylegan
[4] https://github.com/NVlabs/stylegan2-ada-pytorch
[5] https://github.com/openai/guided-diffusion
[6] https://github.com/Rayhane-mamah/Efficient-VDVAE



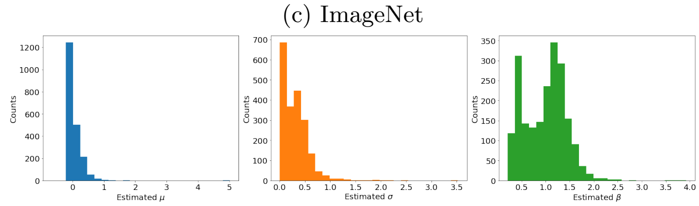
(a) CIFAR10



(b) CelebA



(c) ImageNet



(d) FFHQ

Fig. C: Histograms of the estimated parameters (μ, σ, β) for the 2048 dimensions of the Inception features for images from (a) CIFAR10, (b) CelebA, (c) ImageNet, and (d) FFHQ datasets.

The initial values of the parameters in our method are so determined that fast convergence of the maximum likelihood estimation using (10) is achieved. The initial value of μ is set to the peak location of the histogram of Inception features for each dimension. The initial values of σ and β are empirically set to $1.5\hat{\sigma}$ and 0.67, respectively, where $\hat{\sigma}$ is the sample standard deviation. It takes about 1.5 hours to estimate the distributions of 2048-dimensional Inception features for 50000 images using a 3.7 GHz quad-core Intel Xeon® CPU.

C More Results of Density Estimation

To supplement the results in Section 5.3, Fig. C shows the histograms of the parameters for all feature dimensions.

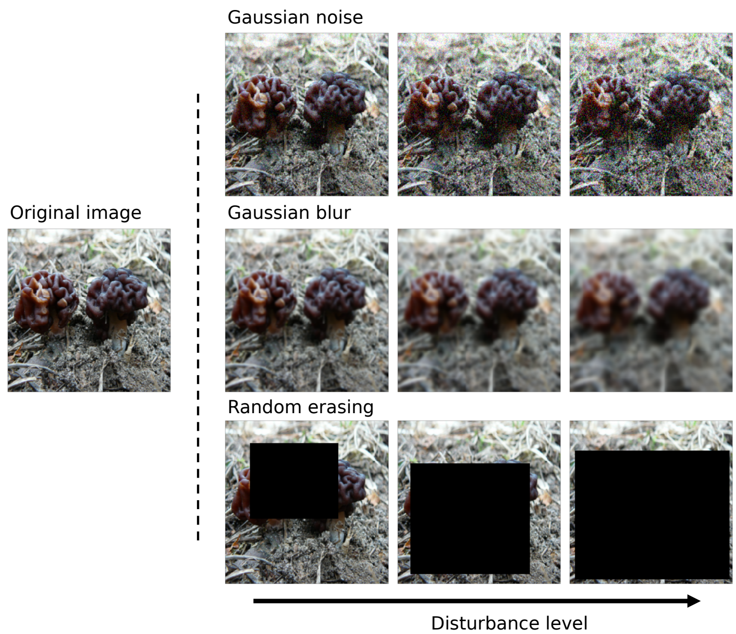


Fig. D: Example of disturbed images.

D Disturbed Images

An example of disturbed images used in Section 5.4 is shown in Fig. D.

References

1. Brock, A., Donahue, J., Simonyan, K.: Large scale GAN training for high fidelity natural image synthesis. In: Proceedings of the International Conference on Learning Representations (2019)
2. Deng, J., Dong, W., Socher, R., Li, L.J., Kai Li, Li Fei-Fei: Imagenet: A large-scale hierarchical image database. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 248–255 (2009)
3. Dhariwal, P., Nichol, A.Q.: Diffusion models beat GANs on image synthesis. In: Advances in Neural Information Processing Systems (2021)
4. Hazami, L., Mama, R., Thurairatnam, R.: Efficient-VDVAE: Less is more. arXiv (2022), arXiv:2203.13751
5. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of GANs for improved quality, stability, and variation. In: Proceedings of the International Conference on Learning Representations (2018)
6. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Jun 2019)

7. Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., Aila, T.: Analyzing and improving the image quality of StyleGAN. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (Jun 2020)
8. Krizhevsky, A., et al.: Learning multiple layers of features from tiny images. Tech. rep. (2009)
9. Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of the International Conference on Computer Vision. pp. 3730–3738 (Dec 2015)
10. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of the International Conference on Learning Representations (2016)