

1 Response of ICML submission 3891

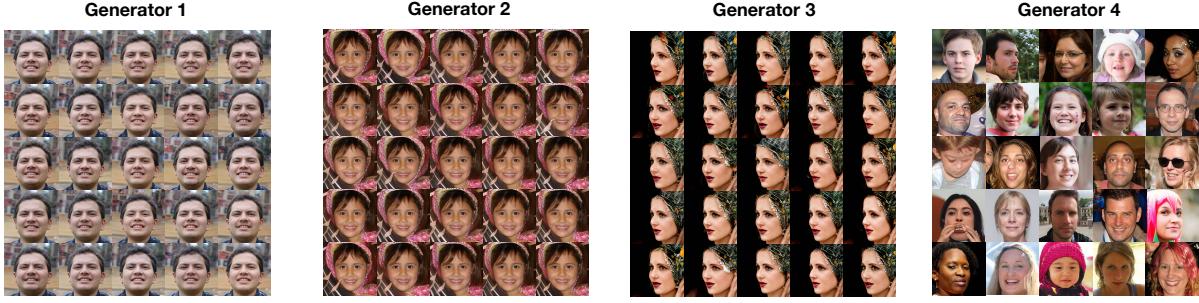


Figure 1: The counterintuitive ranking given by FID-avg score on FFHQ dataset. FID-avg even gives better scores to collapsed models (generators 1 to 3). Here we set truncation $\tau = 0.1$ as we explained in Section 5.3.

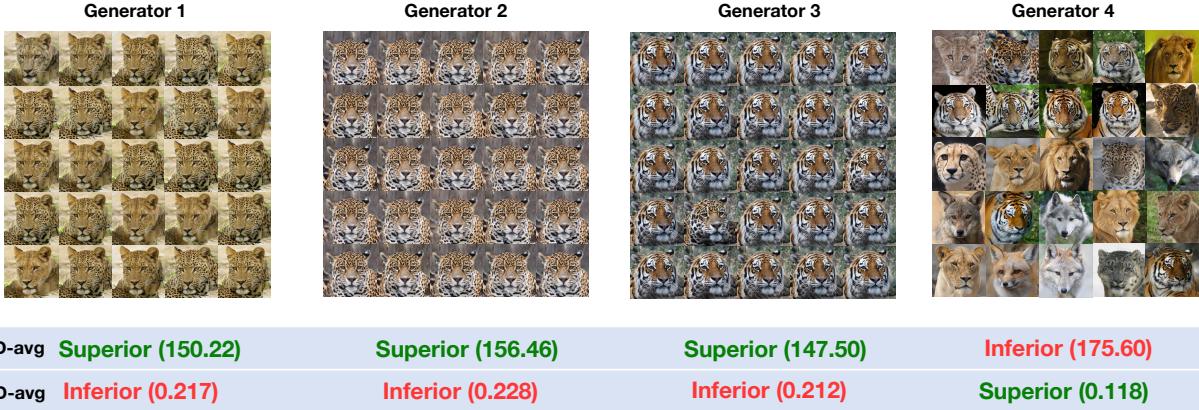


Figure 2: The counterintuitive ranking given by FID-avg score on AFHQ dataset.

Evaluation on Variance-Limited Mixed Federated CIFAR10 Evaluation on Variance-Limited Mixed Federated CIFAR100

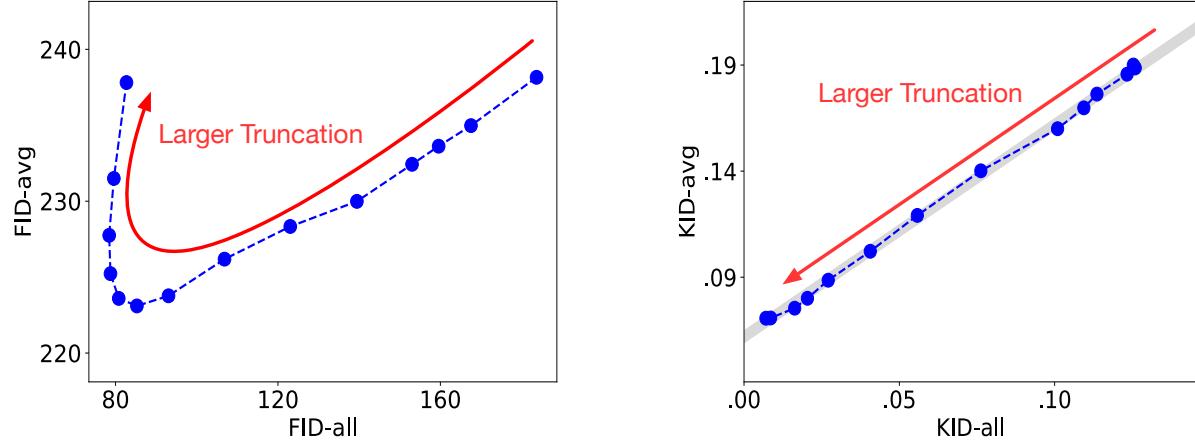


Figure 3: Evaluation results on CIFAR10 where clients hold samples from two classes. The setting is similar to that in Fig. 16 in main text but with sample overlap.