Supplementary Materials for Self-Distilled StyleGAN: Towards Generation from Internet Photos

ANONYMOUS AUTHOR(S)
SUBMISSION ID: 289

ACM Reference Format:
Anonymous Author(s). 2022. Supplementary Materials for Self-Distilled StyleGAN: Towards Generation from Internet Photos. *ACM Trans. Graph.* 1, 1 (January 2022), 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

A IMPLEMENTATION DETAILS

Self-filtering. In our experiments we use our own Tensorflow 2.0 StyleGAN2 implementation with the same hyperparameters as in the official implementation. The encoder architecture is based on the StyleGAN2 discriminator. We train our encoder jointly with the StyleGAN generator using 8 Tesla V100 GPUs for $250k$ steps, with a learning rate of 0.002 for both encoder and generator and batch size of 16. We set $\lambda_w = \lambda_{L1} = \lambda_{\text{LPIPS}} = 0.1$ in all our experiments. We train the generator over the filtered subset for $500K$ iterations in the comparisons and ablations and $1500K$ iterations for the final models.

Multi-Modal Truncation. Our implementation employs the standard Kmeans algorithm, where we sample 60,000 latent codes by passing random noise vectors through StyleGAN's mapping network. We obtain $N = 64$ clusters using Python Sklearn implementation with the default parameters, except for 'init' which we set to random.

B ADDITIONAL ABLATIONS AND COMPARISONS

StyleGAN3. [Karras et al. 2021] has been demonstrated to better address unaligned data. Therefore, we validate that our filtering method is still valuable even when utilizing this architecture. Trained over the challenging LSUN-bicycle, StyleGAN3 achieves an FID score of 6.5 on unfiltered data and 2.6 on our filtered dataset. StyleGAN2 obtains 5.42 and 3.66, respectively. We conclude that StyleGAN3 alone is insufficient to handle challenging uncurated datasets collected from the internet, and also benefits from our filtering scheme. Since it is yet to be shown that the semantic disentanglement of StyleGAN2 is still preserved in StyleGAN3, we base our framework on the irrefutable StyleGAN2 model.

Number of clusters. The number of clusters, N, is affecting the obtained clustering quality in our multi-modal truncation. Insufficient number results in each cluster consisting of multiple modalities, leading to inferior diversity, similar to the global mean truncation. On the other hand, an excessive number of clusters yields centers with lower visual quality resulting in additional artifacts. First, We have used the commonly practiced elbow method to determine the number of clusters. We evaluated the elbow method based on KMeans inertia and FID with a fixed $\psi = 0.5$. Both results in nearly $N = 64$ for most datasets, while increasing N to 128 clusters gains only a minor FID improvement. In addition, we study the preferable number of clusters, by conducting a user study (AMT). We presented side-by-side a triplet of images using 32, 64, or 128 clusters for our proposed truncation approach. The participants were asked to choose the most realistic image. As can be seen in Table 1, our method is not highly sensitive to the exact number of clusters. Yet, rich and diverse domains that consist of many modalities, e.g. dogs, which differ in pose and breed, achieve slightly better results with 128 cluster centers. On the other hand, the score of less diverse domains, such as horses, slightly improves when using only 32 clusters. Overall, we conclude that 64 clusters perform well for all datasets, and therefore, we use this configuration for all our experiments.

Dataset	Number of clusters 32	64	128
Internet Lions	34.3%	33.5%	32.2%
Internet Parrots	34.0%	32.0%	34.0%
LSUN-Horses	34.5%	34.1%	31.4%
LSUN-Bicycles	32.6%	32.3%	35.1%
Internet Dogs	26.4%	36.0%	37.6%

Table 1. **Ablation user study (AMT) results for the number of clusters used in our multi-model truncation.** *We presented side-by-side a triplet of images using* 32, 64, *or* 128 *cluster centers for our proposed truncation scheme. The participants were asked to choose the most realistic image. Overall, we observe that* 64 *clusters perform well for all datasets.*

	Lions	Parrots	Horses	Bicycles	Dogs
Latent Assignment	43.1%	44.5%	41.8%	44.1%	45.9%
Ours	**56.9%**	**55.5%**	**58.2%**	**55.9%**	**54.1%**
Clamping	42.6%	43.5%	30.5%	45.7%	35.7%
Ours	**57.4%**	**56.5%**	**69.5%**	**54.3%**	**64.3%**

Table 2. **Additional ablation and comparison for multi-modal truncation.** *For each comparison, we conduct a user study where the raters requested to choose the more realistic result out of two given images — ours and the evaluated baseline. We report the percentage of raters in favor of each. Top: We replace the LPIPS-based assignment with the latent-based assignment. Bottom: we perform clamping to the global mean [Kynkäänniemi et al. 2019].*

Latent-based cluster assignment ablation. We further validate our design choice, assigning the "nearest" cluster center using the LPIPS

0730-0301/2022/1-ART $15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

perceptual distance rather than using the euclidean distance between the latent codes. We compare the two alternatives by conducting a user study (AMT), where the participants are asked to choose the more realistic image. As shown in Table 2, LPIPS-based assignment indeed outperforms the latent space-based assignment.

Clamping truncation comparison. We compare our multi-modal truncation method to the clamping toward the global mean method [Kynkäänniemi et al. 2019]. Similar to other comparisons and ablations, we evaluate this ablation via an AMT user study. Again, we tune the parameters of the different methods to reach the same FID. As can be seen in Table 2, our results are superior, therefore, we further conclude that using clamping to global mean is also not adequate for challenging multi-modal data.

REFERENCES

Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2021. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems* 34 (2021).

Tuomas Kynkäänniemi, Tero Karras, Samuli Laine, Jaakko Lehtinen, and Timo Aila. 2019. Improved Precision and Recall Metric for Assessing Generative Models. arXiv:1904.06991 [stat.ML]