Quantifying Generative Stability: Mode Collapse Entropy Score for Mode Diversity Evaluation - Supplementary Material

Jens Duym, José Oramas, Ali Anwar
IDLab, University of Antwerp - imec
Sint-Pietersvliet 7, 2000 Antwerp, Belgium
jens.duym, jose.oramas, ali.anwar@uantwerpen.be

A. Architectural Adjustments for a Higher-Resolution DCGAN

To generate 96×96 images instead of 64×64 images, as proposed in [1], we adjusted the architecture of DCGAN.

A.1. Generator

We used five transposed convolutional layers. Each layer uses Batch Normalization and a ReLU activation, except for the final layer, which uses a Tanh activation. The details of each layer are shown in Table 1. For the model trained with a batch size of 4, Instance Normalization replaced Batch Normalization for stability.

A.2. Discriminator

The discriminator consists of five convolutional layers. Each layer uses a LeakyReLU activation and Batch Normalization, except for the final layer, which uses a Sigmoid activation. The details of each layer are presented in Table 2. We also adjusted the Batch Normalization layers to Instance Normalization in the model trained with a batch size of 4.

A.3. Training Configuration

The models were trained using the following hyperparameters:

- **Latent Vector Size** (nz): 128
- **Feature Map Sizes** (ngf, ndf): 64
- **Image Resolution**: 96×96
- **Number of Channels** (nc): 3 (RGB)
- **Batch Size**: 4, 64, and 128
- **Learning Rate** (lr): 0.0002
- **Optimizer**: Adam with $\beta_1 = 0.5$ and $\beta_2 = 0.999$

We stopped training the models after a certain number of epochs to ensure comparable performance regarding the FID score. The scores are 99.81, 92.32, and 93.11 for the models with batch sizes of 4, 64, and 128, respectively. It should be noted that these models perform far from perfectly with respect to image fidelity. However, this work focuses on another quality aspect: mode collapse.

B. Generated Images from our DCGAN Models

Figures 1 to 3 illustrate generated images of our DCGAN model trained with batch sizes of 4, 64, and 128, respectively.

Acknowledgements: This work was supported by the Research Foundation Flanders (FWO) under Grant Number 1S52323N.

References

[1] Alec Radford. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. 1

Table 1. Generator Architecture

Layer	Operation	Kernel Size	Stride	Padding	Activation	BatchNorm
1	ConvTranspose2d ($nz \to ngf \times 8$)	6×6	1	0	ReLU	Yes
2	ConvTranspose2d ($ngf \times 8 \to ngf \times 4$)	4×4	2	1	ReLU	Yes
3	ConvTranspose2d ($ngf \times 4 \to ngf \times 2$)	4×4	2	1	ReLU	Yes
4	ConvTranspose2d ($ngf \times 2 \to ngf$)	4×4	2	1	ReLU	Yes
5	ConvTranspose2d ($ngf \to nc$)	4×4	2	1	Tanh	No

Table 2. Discriminator Architecture

Layer	Operation	Kernel Size	Stride	Padding	Activation	BatchNorm
1	Conv2d ($nc \to ndf$)	4×4	2	1	LeakyReLU(0.2)	No
2	Conv2d ($ndf \to ndf \times 2$)	4×4	2	1	LeakyReLU(0.2)	Yes
3	Conv2d ($ndf \times 2 \to ndf \times 4$)	4×4	2	1	LeakyReLU(0.2)	Yes
4	Conv2d ($ndf \times 4 \to ndf \times 8$)	4×4	2	1	LeakyReLU(0.2)	Yes
5	Conv2d ($ndf \times 8 \to 1$)	6×6	1	0	Sigmoid	No

Figure 1. Generated images of our DCGAN (batch size: 4, FID: 99.81, MCE-P: 0.0955, MCE-C: 0.0005).

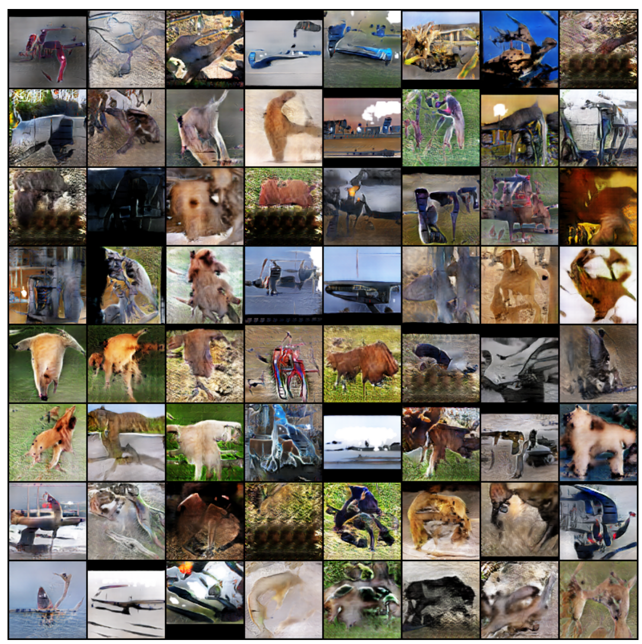

Figure 2. Generated images of our DCGAN (batch size: 64, FID: 92.32, MCE-P: 0.2001, MCE-C: 0.0142).

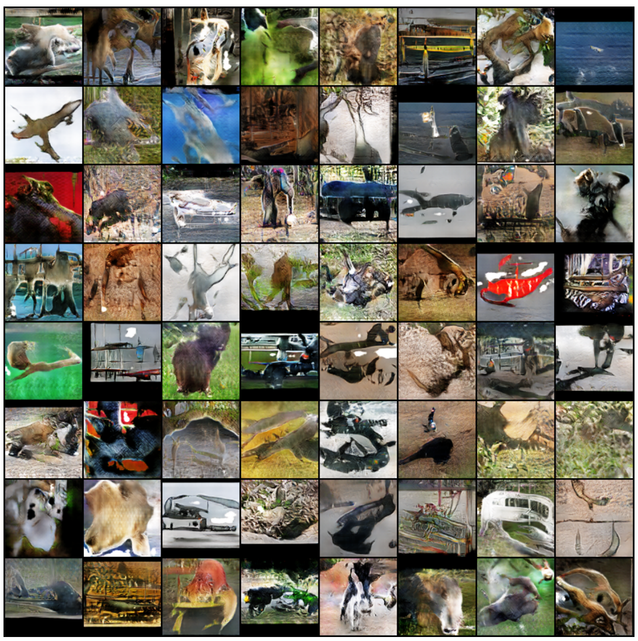

Figure 3. Generated images of our DCGAN (batch size: 128, FID: 93.11, MCE-P: 0.2280, MCE-C: 0.0336).