Dog Image Generation Competition on Kaggle

Wendy Kan
Kaggle / Google
wendykan@google.com

Phil Culliton
Kaggle / Google
philculliton@google.com

Douglas Sterling
Kaggle / Google
dster@google.com

Abstract

We present a novel format of machine learning competitions where a user submits code that generates images trained on training samples, the code then runs on Kaggle, produces dog images, and user receives scores for the performance of their generative content based on 1. quality of images, 2. diversity of images, and 3. memorization penalty. This style of competition targets the usage of Generative Adversarial Networks (GAN)[4], but is open for all generative models. Our implementation addresses overfitting by incorporating two different pre-trained neural networks, as well as two separate "ground truth" image datasets, for the public and private leaderboards. We also have an enclosed compute environment to prevent submissions of non-generated images. In this paper, we describe both the algorithmic and system design of our competition, as well as sharing our lessons learned from running this competition [6] in July 2019 with 900+ teams participating and over 37,000 submissions and their code received.

1 Design Overview

In this competition, we created an enclosed compute environment in Kaggle Kernels, where users are restricted to a Docker environment with Python/R with common ML libraries; but without internet access or any other external datasets other than the training images, or any pre-trained models. Users have access to GPUs (NVIDIA P100) and up to 9 hours of compute time and 16 GB of RAM. This setup ensures the integrity of the competition, such that no external images can be directly submitted (altered or not). It also ensures that the generative models can be trained in a fixed amount of time and resources.

There are many ways to evaluate GANs [1]. We present MiFID (Memorization-informed Fréchet Inception Distance), which is a modification from Fréchet Inception Distance (FID)[5] with addition of a novel memorization distance. See the details and mathematical representation here and code here. Figure 1 shows our workflow for evaluating the submissions. Two different NNs are applied separately to produce the public and private leaderboards, as well as two separate dog datasets.

2 Results

We had 924 teams joining the competition which produced over 37,000 submissions over the course of 8 weeks. The highlight of the results include: BigGAN [2] is used in 4 of the top 5 winners, where the Kaggle community's ability to train BigGAN from scratch in 9 hours is impressive.

Memorization GANs [3], where the network is designed specifically to memorize and generate training images, are very difficult to detect using only the submitted images. Our memorization distance was very effective in penalizing memorization, however, it requires a careful decision of the cosine distance threshold ϵ, which was determined by manually inspecting users' submitted code.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

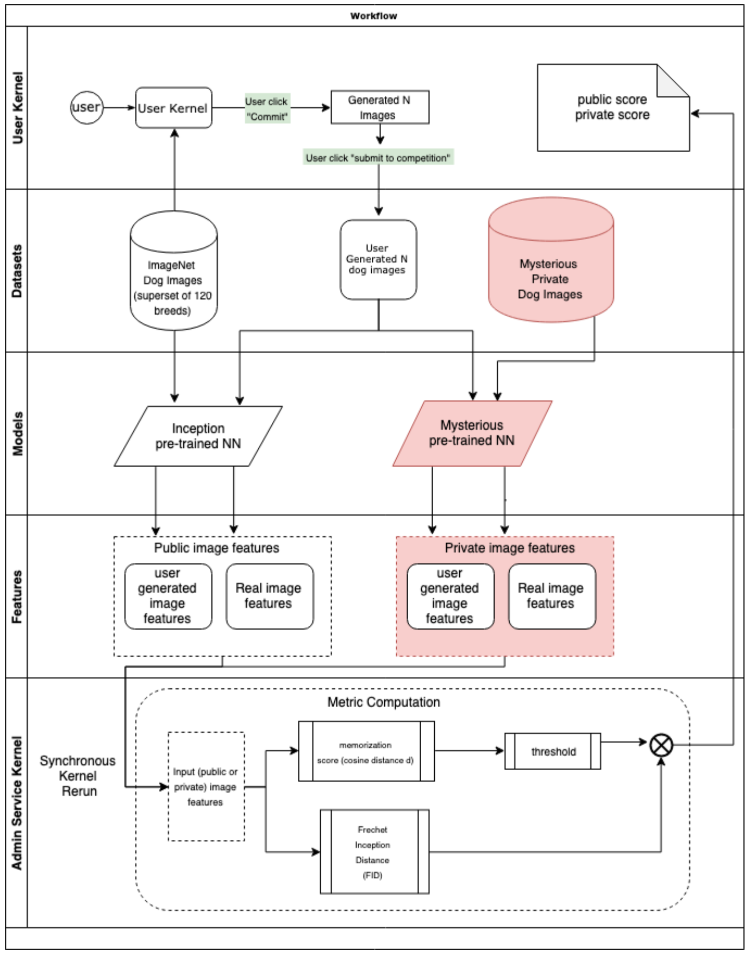

Figure 1: Kaggle's workflow calculating MiFID for public and private scores

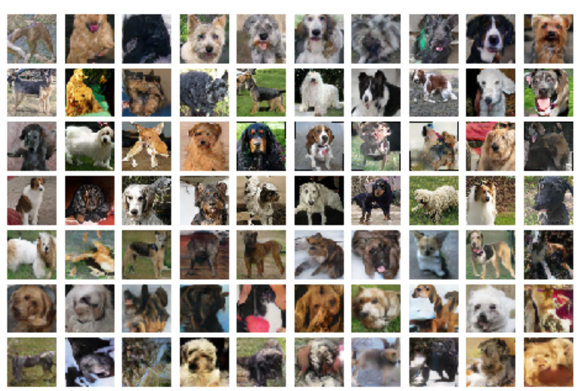

Figure 2: Submissions from ranks 1 (first row), 2, 3, 5, 10, 50, 100 (last row) on the private leaderboard. Each row is a random sample of 10 images from the same team. Visually you can see the quality of the generated images getting lower as the ranks get higher.

Acknowledgments

Colin Raffel, Google Brain: consulting on system design and metric evaluation workflow.

Andrew Brock, DeepMind: consulting on system design and metric evaluation workflow.

Julia Elliott, Kaggle/Google: managing all the correspondence and ensuring competition runs smoothly.

Ching-yuan (Andrew) Bai, National Taiwan University: analyzing results, thoroughly understanding of memorization GANs, and inspection of submitted code.

References

[1] Ali Borji. Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding*, 179:41–65, 2019.

[2] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*, 2018.

[3] Chris Deotte. *Dog Memorizer GAN*, 2019. https://www.kaggle.com/cdeotte/dog-memorizer-gan.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[5] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*, pages 6626–6637, 2017.

[6] Wendy Kan. *Generative Dog Images: Experiment with creating puppy pics*, 2019. https://www.kaggle.com/c/generative-dog-images/.