Generative Adversarial Metric

Daniel Jiwoong Im[1], Chris Dongjoo Kim[2], Hui Jang[2], and Roland Memisevic[1]
[1]University de Montreal and [2]York University

Introduction

- A problem with generative adversarial models is that there is not a clear way to evaluate them quantitatively.

  In the past,

  - [2] evaluated GANs by looking at the single nearest-neighbour data from the generated samples.
  - [1] evaluated based on human inspections. In that case, the discriminator can be viewed as a human, while the generator is a trained GAN.
  - [6] evaluated based on classification performance.

  We propose generative adversarial metric which compares two GANs by having them engage in a "battle" against each other.

Background

Generative Adversarial Networks (GAN) consists of generative and discriminative model, G and D.

The generative model generates samples that are hard for the discriminator D to distinguish from real data.

The discriminative model tries to avoid getting fooled by the generative model G.

Trained by playing *a minmax game* as follows:

$$\min_{\theta_G} \max_{\theta_D} V(D,G) = \min_G \max_D \Big[\mathbb{E}_{\boldsymbol{x}\sim p_{\mathcal{D}}}\big[\log D(\boldsymbol{x})\big] + \mathbb{E}_{\boldsymbol{z}\sim p_{\mathcal{G}}}\big[\log\big(1-D(G(\boldsymbol{z}))\big)\big]\Big]. \quad (1)$$

where θ_G and θ_D are the parameters of discriminator and generator, respectively.

Method

IDEA: Since every generative adversarial models consists of a discriminator and a generator in pairs, we can exchange the pairs and have them play the generative adversarial game against each other.

Consider two generative adversarial models, M_1 and M_2. Each model consists of a generator and a discriminator,

$$M_1 = \{(G_1, D_1)\} \text{ and } M_2 = \{(G_2, D_2)\}. \quad (2)$$

In the training stage, both models are being trained to prepare them for the battle with one another. In the test phase, model M_1 plays against model M_2 by having G_1 try to fool D_2 and vice-versa.

	M_1	M_2
M_1	$D_1(G_1(\boldsymbol{z}))\,,\, D_1(\boldsymbol{x}_{train})$	$D_1(G_1(\boldsymbol{z}))\,,\, D_1(\boldsymbol{x}_{test})$
M_2	$D_2(G_2(\boldsymbol{z}))\,,\, D_2(\boldsymbol{x}_{test})$	$D_2(G_2(\boldsymbol{z}))\,,\, D_2(\boldsymbol{x}_{train})$

We look at the following ratios between the discriminative scores of the two models:

$$r_{test} = \frac{\epsilon\big(D_1(\boldsymbol{x}_{test})\big)}{\epsilon\big(D_2(\boldsymbol{x}_{test})\big)} \text{ and} \quad (3)$$

$$r_{samples} = \frac{\epsilon\big(D_1(G_2(\boldsymbol{z}))\big)}{\epsilon\big(D_2(G_1(\boldsymbol{z}))\big)}, \quad (4)$$

where $\epsilon(\cdot)$ outputs the classification error rate. These ratios allow us to compare the model performance.

In order to address this issue, our proposed evaluation metric qualifies the sample ratio to be judged by the test ratio as follows: winner =

$$\begin{cases} \text{M1} & \text{if } r_{sample} < 1 \text{ and } r_{test} \simeq 1 \\ \text{M2} & \text{if } r_{sample} > 1 \text{ and } r_{test} \simeq 1 \\ \text{Not Applicable} & \text{otherwise} \end{cases} \quad (5)$$

Our proposed evaluation metric qualifies the sample ratio using the test ratio by defining the winning model in Equation 13. For more details, please refer to our paper.

GAM is able to compare other models by observing the error rate of GAN's discriminators based on the samples of other generative model.

Figure

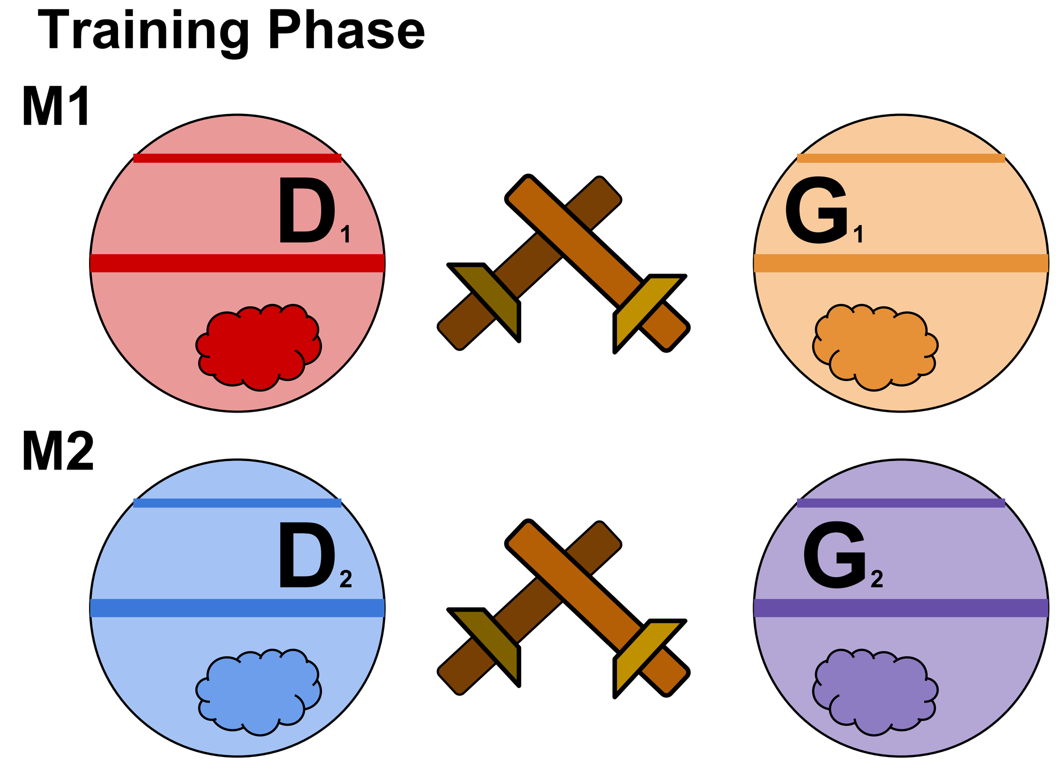

Figure 1: The top left scene mask is the ground truth of the IP6 dataset, and the next six images are the predicted scene masks using various types of SSNNs trained with 30 labeled examples.

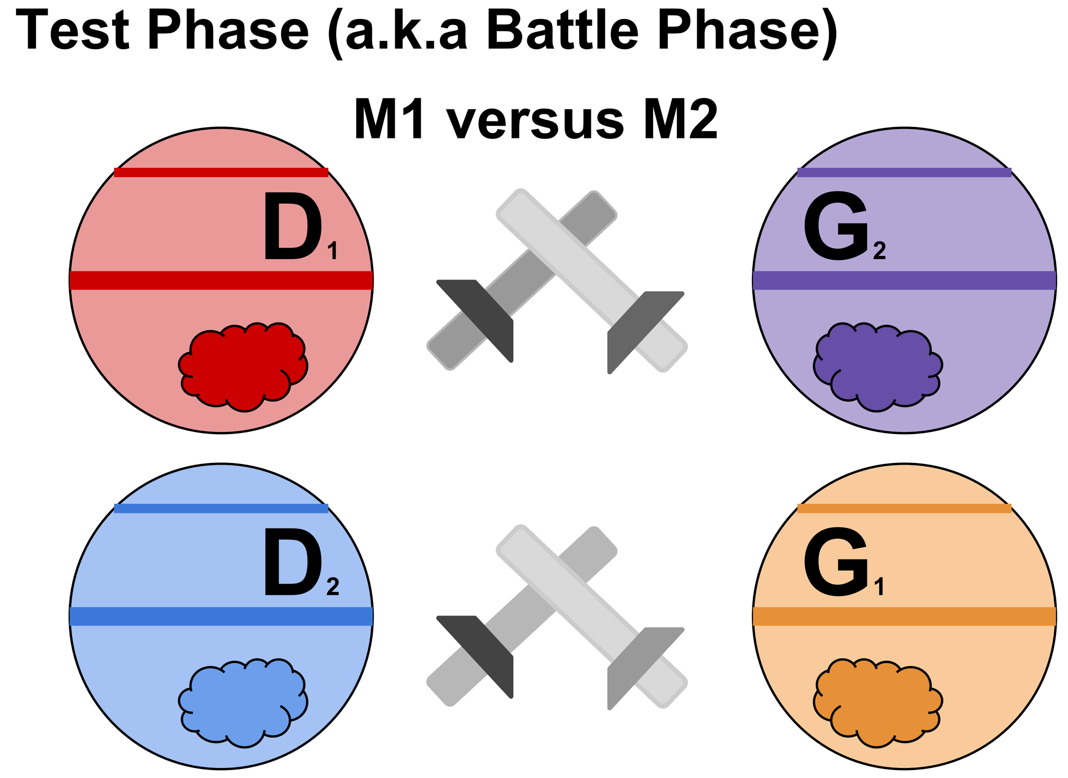

Figure 2: The top left scene mask is the ground truth of PaviaU, and the next six images are the predicted scene masks using various types of SSNNs trained with 30 labeled examples.

Experiments

We considered model two types of model GAN and GRAN. GRAN is generative adversarial neural network with the recurrent connections on generator of the model.

The performance of GAN models based on GAM metric is presented in the below Table. Note that GRAN1 is refered to as GAN.

The performance of GAN models versus non GRAN based on GAM metric is presented in the below Table.

Battler	Error
GAN[2] vs. DVAE[4]	0.058
GRAN3[5] vs. DVAE[4]	0.01
GAN[2] vs. DRAW[3]	0.347
GRAN3[5] vs. DRAW[3]	0.106

Data set	Battler	Test Ratio	Sample Ratio	Winner
MNIST	GAN vs. GRAN3	0.79	1.75	GRAN3
	GAN vs. GRAN5	0.95	1.19	GRAN5
CIFAR10	GAN vs. GRAN3	1.28	1.001	GRAN3
	GAN vs. GRAN5	1.29	1.011	GRAN5
	GRAN3 vs. GRAN5	1.00	2.289	GRAN5
LSUN	GAN VS. GRAN3	0.95	13.68	GRAN3
	GAN vs. GRAN5	0.99	13.97	GRAN5
	GRAN3 vs. GRAN5	0.99	2.38	GRAN5

References

[1] Denton E., Chintala S., Szlam A, Fergus, R., Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks, NIPS 2015
[2] Goodfellow I. J., Pouget-Abadie J., Mirza M. Xu B., Warde-Farley D., OzairâĂă S., Courville A., Bengio, Y. Generative Adversarial Nets, NIPS, 2014
[3] Gregor K., Danihelka I., Graves A., Rezende D., and Wierstra D,. Draw: A recur- rent neural network for image generation. ICML 2015
[4] Im D., Ahn S., Memisevic R., Bengio, Y. Denosing Criterion for Variational Auto-encoding Framework, http://arxiv.org/abs/1511.06406, 2016
[5] Im D., Kim C., Jang H., Memisevic R., Generating Images with Recurrrent Adversarial Networks, http://arxiv.org/abs/1602.05110, 2016
[6] Radford A., Metz L., Chintala S., Unsupervised Representation Learning With Deep Convolutional Generative Adversarial Networks, ICLR 2015