General Adversarial Networks (GANs)

Lauren McGinney, Chukwudi Onyema Ajoku, Galina Kulya and Anslem Ordia

Department of Applied Artificial Intelligence, Teesside University, United Kingdom

Objectives

To demonstrate the knowledge and skills gained during this module, the output of two different GANs - Fully Connected GAN (FCGAN) and Deep Convolutional GAN (DCGAN) - will be compared on 4 standard machine learning datasets:

- MNIST
- Fashion MNIST
- CIFAR10
- CIFAR100

The quality of generated images will be presented and evaluated.

Dataset

The datasets are high quality, curated, commonly used for machine learning research, cited in peer-reviewed academic journals and can easily be loaded from the Keras library. Labels are available, but are not required to train GANs and generate images. One limitation is the low resolution of images, resulting in low quality generated images. Alternatively, many image datasets are available from websites like www.kaggle.com/datasets, to use in future GAN applications.

Methodology

The Keras library was used in Python to design, train and visualise the models and results in Google Colab. Available **here.** FCGAN models had 4 dense layers, while DCGAN models had 4 convolutional layers. Generator models used tahn activation, while discriminator models used sigmoid. Variables common to all architectures:

- epochs = 10, (batch size = 16, steps per epoch = 3750)
- optimiser = Adam(0.0002, 0.5)
- loss = "binary_crossentropy"
- activation after every layer = LeakyReLU(0.2)

Generated images were printed after each epoch. After training, D and G loss were plotted.

Model

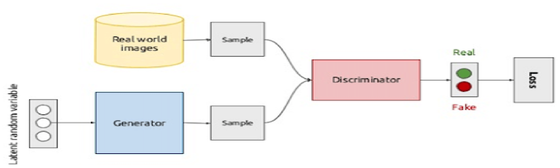

Figure 1:Standard GAN Architecture

Theory

GAN is based on Game Theory. It converges when the discriminator and the generator reach a **Nash equilibrium**. The equation for this is given as:

$$\min_G \max_D V(D,G) =$$

$$\mathbb{E}_{x\sim p_{data}(x)}[logD(x)] + \mathbb{E}_{z\sim p_z(z)}[log(1-D(G(z)))]$$

where D and G represent the Discriminator and the Generator respectively, z stands for random vector for the generator and $D(x)$ is the chances that an output is real.

Result and Evaluation

Figure 2:A Comparison of FC vs DC GAN outputs

Evaluation Metrics

Frechet-Inception Distance (FID) was the metrics used for evaluation. It captures the similarity of generated images to real ones better than the Inception Score. The lower the FID, the better the image generated. (Heusel et al. 2017). Its equation is given as:

$$FID = \| \mu_r - \mu_g \|^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{\frac{1}{2}})$$

where $\mathcal{X}_r \sim \mathcal{N}(\mu_r, \Sigma_g)$ and $\mathcal{X}_g \sim \mathcal{N}(\mu_g, \Sigma_g)$ are the dimensional activations of the Inception-v3 pool3 layer for real and generated samples respectively.

Observation

These examples are randomly sampled, and show:

- All GANs produced legible images of good quality.
- FCGAN generated images are more grainy, and the class is not as clear as DCGAN generated images.
- CIFAR generated images are very similar to the real dataset - objects, but not all details can be seen.
- Similar quality to other available results (Theiler 2021).

Observation cntd.

	MNIST	FASHION_MNIST	CIFAR10	CIFAR100
FC-GAN	2.74	4.07	7.83	14.07
DC_GAN	1.57	38.06	3.20	5.38

- The DC-GAN FID score for the MNIST_FASHION dataset did not behave as expected. The scores are high but on the contrary, images produced by DC-GAN are much more superior.
- FID may still not be the most perfect way to evaluate GANs

Challenges and Recomendations

- There is need for faster GPU as google colab always time out when time slot is used up. GANs are computationally expensive.
- There is no objective loss function used to train the GAN generator models and no way to objectively assess the progress of the training and the relative or absolute quality of the model from loss alone. Models must be evaluated using manual inspection of the quality of the generated synthetic images.

Relevant risks, professional, social, ethical, security and privacy issues

Deep learning based software can cause threats to privacy, democracy and national security.

- Anything is possible. Videos, images and audios might no longer serve as concrete evidence in court.
- Can be weaponized to harass and humiliate (eg, fake porn with celebrity face)

Conclusion

- The architectures and model parameters used worked well.
- DCGANs outperformed FCGANs, but all models produced recognisable images.
- Future work could test the design on other datasets, investigate fine-tuning the models or other GAN architectures (e.g. StyleGAN, CycleGAN) and quantitative evaluation of results.
- There is need to guard against the misuse of GANs. Policies can be made to this effect.

References

Goodfellow, Ian J. et al. (2014). *Generative Adversarial Networks*. arXiv: 1406.2661 [stat.ML].

Heusel, Martin et al. (2017). "Gans trained by a two time-scale update rule converge to a local nash equilibrium". In: *arXiv preprint arXiv:1706.08500*.

Radford, Alec, Luke Metz, and Soumith Chintala (2016). *Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks*. arXiv: 1511.06434 [cs.LG].

Salimans, Tim et al. (2016). *Improved Techniques for Training GANs*. arXiv: 1606.03498 [cs.LG].

Theiler, Sebastian (2021). *Implementing A GAN in Keras*. URL: https://medium.com/analytics-vidhya/implementing-a-gan-in-keras-d6c36bc6ab5f.