Evaluation of Spectral Normalization for GANs Using Inception Score

Eysteinn Gunnlaugsson, Egill Vignisson, Charles Hamesse

Group 16

Abstract. We experiment various Deep Convolutional Generative Adversarial Networks (DCGANs) for image generation. We start with a vanilla DCGAN and gradually add features that are expected to improve learning in terms of speed, stability and quality of generation. We evaluate our models using the inception score and discuss its relevance on cifar10 and a dataset that we collected ourselves consisting of reptile images. Finally, we discuss our findings and challenges and draw conclusions on how the features we implemented help the training of GANs.

Keywords: Generative models, generative adversarial networks, image generation

1 Introduction

Motivate the problem you are trying to solve, attempt to make an intuitive description of the problem and also formally define the problem. (1-2 pages including title, authors and abstract)

The purpose of this project is to investigate the performance of the different types of Generative Adversarial Networks (GANs) [1] for image generation as well as possible improvement options. The project was originally defined based on varying levels of priority where everything assigned priority 1 was promised to be completed.

– Implement Deep Convolutional Generative Adversial Network with original loss [2] (priority 1)
– Implement the inception score metric citesalimans2016improved (priority 1)
– Implement Spectral Normalization (priority 1)
– Evaluate all our GANs on our reptile dataset (priority 1)
– Implement other losses (LSGAN, WGAN) [3] (priority 2)
– Evaluate GANs on CIFAR-100 (priority 2)
– Implement mini-batch discrimination and or other improvements [4]. (priority 3)

In order to evaluate the performance of these GANs, the evaluation metric known as the inception score, as described in [4], will be implemented. Furthermore a data set consisting of roughly 30K animal pictures (mostly reptiles), was fetched from the Flickr API while other well known options such as a subset of CIFAR-100 or ImageNet were thought of as possible replacements in the case of unsatisfactory results.

2 Background

We present the framework of GANs

2.1 Generative Adversarial Networks

The learning process of the GANs is to train a discriminator D and a generator G simultaneously. The target of G is to learn the distribution p_g over data $\boldsymbol{x}$. G starts from sampling input variables $\boldsymbol{z}$ from a uniform or Gaussian distribution $p_z(\boldsymbol{z})$, then maps the input variables $\boldsymbol{z}$ to data space $G(\boldsymbol{z};\theta_g)$ through a differentiable network. On the other hand, D is a classifier $D(\boldsymbol{x};\theta_d)$ that aims to recognize whether an image is from training data or from G. The minimax objective for GANs can be formulated as follows:

$$\min_G \max_D V_{\text{GAN}}(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]. \tag{1}$$

2.2 Inception Score

2.3 Spectral Normalization

2.4 Least Squares Generative Adversarial Networks

Viewing the discriminator as a classifier, regular GANs adopt the sigmoid cross entropy loss function. As stated in Section **??**, when updating the generator, this loss function will cause the problem of vanishing gradients for the samples that are on the correct side of the decision boundary, but are still far from the real data. To remedy this problem, we propose the Least Squares Generative Adversarial Networks (LSGANs). Suppose we use the a-b coding scheme for the discriminator, where a and b are the labels for fake data and real data, respectively. Then the objective functions for LSGANs can be defined as follows:

$$\begin{aligned}\min_D V_{\text{LSGAN}}(D) =& \frac{1}{2}\mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}\big[(D(\boldsymbol{x})-b)^2\big] + \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\big[(D(G(\boldsymbol{z}))-a)^2\big] \\ \min_G V_{\text{LSGAN}}(G) =& \frac{1}{2}\mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}\big[(D(G(\boldsymbol{z}))-c)^2\big],\end{aligned} \tag{2}$$

where c denotes the value that G wants D to believe for fake data.

Benefits of LSGANs The benefits of LSGANs can be derived from two aspects. First, unlike regular GANs, which cause almost no loss for samples that lie in a long way on the correct side of the decision boundary (Figure **??**(b)), LSGANs will penalize those samples even though they are correctly classified (Figure **??**(c)). When we update the generator, the parameters of the discriminator are fixed, i.e., the decision boundary is fixed. As a result, the penalization will make the generator to generate samples toward the decision boundary. On the other hand, the decision boundary should go across the manifold of real data for a successful GANs learning. Otherwise, the learning process will be saturated. Thus moving the generated samples toward the decision boundary leads to making them be closer to the manifold of real data.

Second, penalizing the samples lying a long way to the decision boundary can generate more gradients when updating the generator, which in turn relieves the problem of vanishing gradients. This allows LSGANs to perform more stable during the learning process. This benefit can also be derived from another perspective: as shown in Figure **??**, the least squares loss function is flat only at one point, while the sigmoid cross entropy loss function will saturate when x is relatively large.

3 Approach

Describe the final approach you are take for this problem. For instance, here you would describe the details of the network's architecture. What training parameters and techniques you have used. The computational complexity of your model. And similar questions. To help explain your approach please make figures to accompany your text description. (1-3 pages)

3.1 Data Collection

Images were collected using Flickr's API. Due to varying amounts of quality pictures of objects that would be interesting to base the image generation on, a collection that included mostly reptiles with a fair amount of arachnid's thrown into the mix was ultimately settled upon. In total the data set is made up of approximately 30k color images all re-sized to dimensions of 108x108x3 before training was conducted.

3.2 Frame of Reference

In order to measure the performance of each implementation a frame of reference had to be established. As the golden standard, that every implementation would be compared to, the inception score of the actual data set was calculated resulting in a score of $\mathbf{1.5295098 \pm 0.06244182}$.

4 Experiments

In this section, you should present the results you achieved with various experiments. The results can be presented in tables, plots, etc.

The purpose of the project is to analyze the nature and effectiveness of different GAN architectures as well as different improvement options. The different models but in order to do so a frame of reference was needed

4.1 Inception Score Considerations

We present a challenge related to the computation and evaluation of the inception score. Most authors evaluate the inception score on 50K GAN-generated images, as recommend the authors of the original paper [4]. By running a few preliminary experiments, we quickly realize that on top of the actual training of the network, sampling and computing the inception score are also resource-intensive tasks, and sampling 50K images is simply not possible with the time or resources available for this project.

Now, the number of images considered for evaluating the inception score has an impact on this score, as shows Table 1. This is due to the fact that the inception score not only evaluates the content of a given image but also the distribution of categories amongst the whole set of images resulting from the split. In other words, the score is sensitive to the number of images divided by the number of splits.

Images	Splits	Inception score
256	5	8.13 +- 0.41
512	5	8.04 +- 0.54
1024	5	9.79 +- 0.36

Images	Splits	Inception score
256	10	6.72 +- 0.55
512	10	7.92 +- 0.56
1024	10	8.95 +- 0.44

Table 1: Inception score for various number of samples of the cifar10 dataset.

We choose to stick with 1024 generated images and 5 splits for all of our experiments. With this configuration, we have a target inception score of 9.79. As expected, this is below the claimed inception score of the whole cifar10 dataset, 11.24 [4]. Thus we won't reach state of the art results in terms of inception score, but this isn't an issue since our purpose is to compare various improvements of GAN networks, which isn't affected by this choice. Other considerations on the inception score are explained in [5].

4.2 DCGAN

In order to establish a frame of reference a plain vanilla version of DCGAN using (Fig. 3 shows an example).

Unfortunately we can hardly interpret the losses of GANs, since the generator and discriminator are in a situation of competition where an improvement on the one leads to a deterioration on the other. Consequently, there exists a number of approaches to try and make these networks reach an equilibrium quicker. **Todo:** insert citations or examples if not already done in background theory. However, this aspect of the art of training GANs goes below the scope of our work.

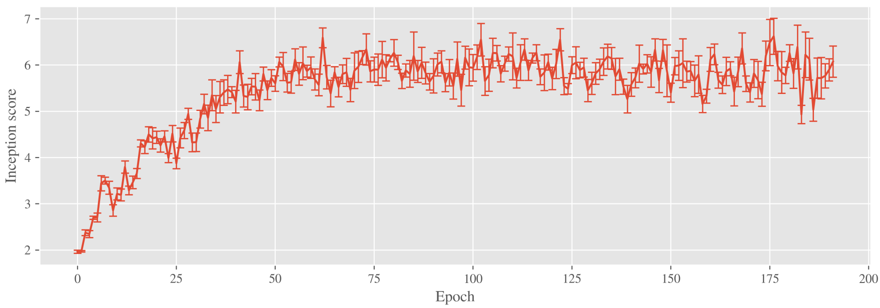

Fig. 1: Inception score over 190 epochs with Vanilla DCGAN.

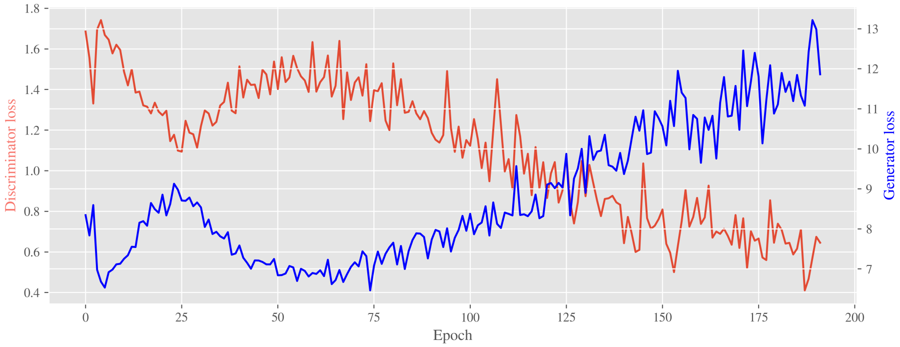

Fig. 2: Losses over 190 epochs with Vanilla DCGAN.

4.3 DCGAN with Spectral Normalization

4.4 LSGAN

4.5 WGAN

5 Conclusions

Explain what conclusions you can draw from these set of experiments? The set of experiments and results reported here should justify some of the design choices described in the previous sections. (3-6 pages)

References

1. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Advances in neural information processing systems. (2014) 2672–2680
2. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR **abs/1511.06434** (2015)
3. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Smolley, S.P.: Least squares generative adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV), IEEE (2017) 2813–2821
4. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Advances in Neural Information Processing Systems. (2016) 2234–2242
5. Barratt, S., Sharma, R.: A note on the inception score. arXiv preprint arXiv:1801.01973 (2018)

A Template examples

Table 2: Font sizes of headings. Table captions should always be positioned *above* the tables. The final sentence of a table caption should end without a full stop

Heading level	Example	Font size and style
Title (centered)	**Lecture Notes ...**	14 point, bold
1st-level heading	**1 Introduction**	12 point, bold
2nd-level heading	**2.1 Printing Area**	10 point, bold
3rd-level heading	**Headings.** Text follows ...	10 point, bold
4th-level heading	*Remark.* Text follows ...	10 point, italic

(Fig. 3 shows an example).

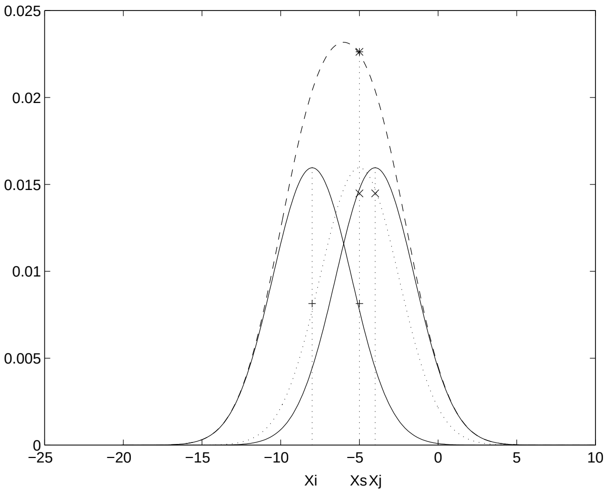

Fig. 3: wuzup

$$\psi(u) = \int_0^T \left[\frac{1}{2}\left(\Lambda_0^{-1}u, u\right) + N^*(-u)\right] dt \tag{3}$$

$$= 0? \tag{4}$$

Please punctuate a displayed equation in the same way as ordinary text but with a small space before the end punctuation.