FLGAN: Focal Loss Generative Adversarial Networks

Ziyue Yang
Department of Computer Science
University of Toronto
Toronto, ON M5S 1A1
ziyue.yang@mail.utoronto.ca
1004804759

Anny Dai
Department of Computer Science
University of Toronto
Toronto, ON M5S 1A1
anny.dai@mail.utoronto.ca
1004881933

Ruyi Qu
Department of Mathematics
University of Toronto
Toronto, ON M5S 2E4
ruyi.qu@mail.utoronto.ca
1004849569

Abstract

Generative Adversarial Networks (GANs), described as *the most interesting idea in the last ten years in machine learning* by the Godfather of AI, Yann LeCun, is a framework that devotes itself into generating natural images, videos, and voices through training generative and discriminative models simultaneously.[1] Though it has achieved great success in recent years, several active concerns, including noisy output and slow convergence speed (or even failure in convergence), have been pointed out in nowadays literature challenging the efficiency of GAN, with the former being addressed by Deep Convolutional Generative Adversarial Network (DCGAN) in 2016.[2] To address the latter, in this paper, we proposed the novel **Focal Loss Generative Adversarial Network (FLGAN)**, an improved version of DCGAN using focal loss to speed up the training process and advance model performance. We evaluate our model on multiple datasets (MNIST, Fashion MNIST, FER-2013) and demonstrated that our proposed method outperforms the DCGAN in image quality generated and efficiency.[3][4][5]

Keywords: Generative Adversarial Network, GAN, DCGAN, Focal Loss, MNIST, Fashion MNIST, FER-2013.

1. Ian Goodfellow et al., "Generative adversarial nets," *Advances in neural information processing systems* 27 (2014).
2. Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015,
3. Li Deng, "The mnist database of handwritten digit images for machine learning research," *IEEE Signal Processing Magazine* 29, no. 6 (2012): 141–142.
4. Han Xiao, Kashif Rasul, and Roland Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms," *arXiv preprint arXiv:1708.07747*, 2017,
5. Ian J Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests," in *International conference on neural information processing* (Springer, 2013), 117–124.

Preprint. Under review.

1 Introduction

"You can have data without information, but you cannot have information without data". This quote from Daniel Keys Moran reveals the importance of data to machine learning and the field of Artificial Intelligence – there is no meaningful training without appropriate data size. However, a lack of data can be common in a real-life scenario. As a response, in 2014, computer theorist Ian Goodfellow and his colleagues proposed a data augmentation technique, named Generative Adversarial Network (GAN), which generates new data samples that are realistic enough to fool a discriminator into believing its realness. Ideally speaking, one can apply GAN to any sort of limited data and then generate NEW meaningful data for the sake of better training outcomes.

However, there are several problems with GAN – images generated with GAN are often quite noisy with slow convergence rate. To address the former issue, Alec Radford et al. proposed the Deep Convolutional Generative Adversarial Network (DCGAN) replacing the multilayer perceptron used by regular GAN with convolutional neural networks (CNN), which in return produces better (less noisy) images. However, the slow convergence rate has not been addressed by DCGAN or any other recent literature.

In this paper, we proposed a improved version of DCGAN, **Focal Loss Generative Adversarial Network (FLGAN)**, by introducing focal loss to penalize low-confidence prediction to speed up convergence time while further improve the quality of generated images. The main contribution of our proposed model are:

- We cut the convergence time of the model needed nearly in half, which significantly helps with the efficiency of the model. Experiments and results can be confidently considered to be plausible as it was trained over three different credible datasets, with the same results showing up.

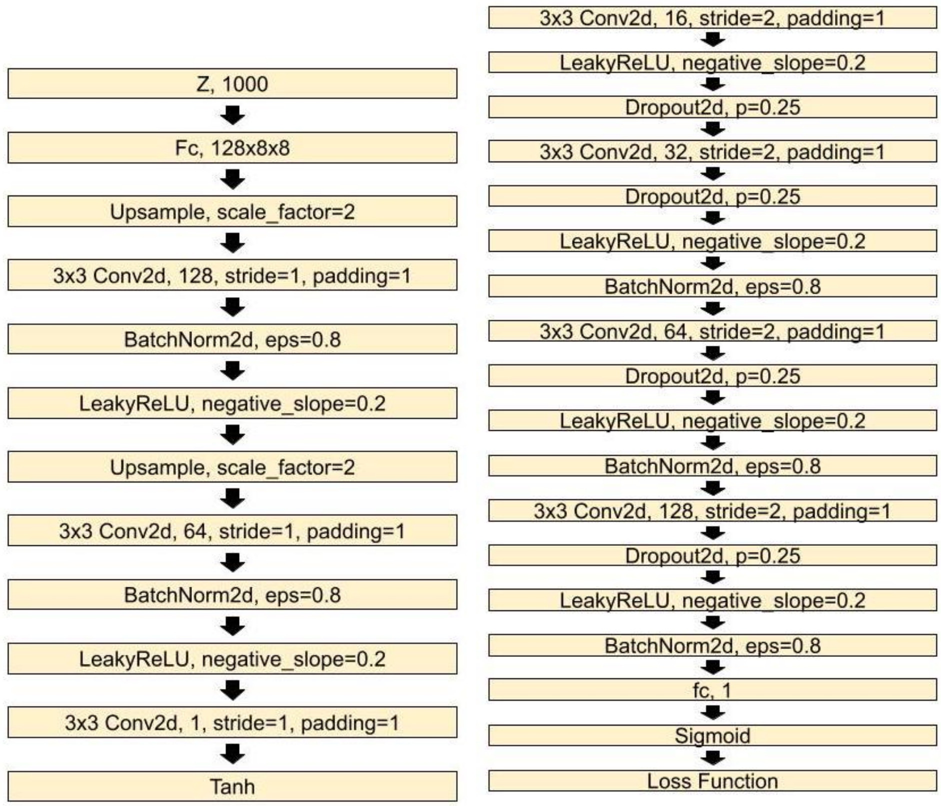
- The source code will be publicly available for replication purposes[6]. Results can be reproduced and used as a benchmark for future studies in image generation.

The rest of this paper is organized as follows: we will be reviewing reviews recent Literature related to image generation, more specially, GAN and DCGAN in section 2 and proceed into explaining our proposed model, choice of dataset and loss function in section 3. In section 4, several experiments are conducted to evaluate the effectiveness of FLGAN. Finally, in section 5 we conclude the paper and provide future direction.

2 Related Work

There exists multiple types of Generative Adversarial Networks (GANs) to generate high-quality images.

The most naive and vanilla approach is to train the GAN initally proposed by Goodfellow.[7] In this approach, we train a generative model G and a discriminative model D simultaneously, where G inputs random noise drawn from some tractable distributions (e.g. Gaussian) and generate fake images, while G attempts to figure out whether a data is from real data or generated from G by computing the probability. The vanilla GAN can learn to generate high-quality samples from the data distribution. Radford *et. al.*[8] proposed the Deep Convolutional Generative Adversarial Networks, an network architecture which combined the convolutional neural networks (CNNs) and GANs, which works better on unsupervised learning tasks.



(a) The Generator of FLGAN (b) The Discriminator of FLGAN

Figure 1: Model Architectures of FLGAN

3 Method and Algorithm

3.1 Model

The model we choose was implemented and modified by Erik Linder-Norén. The original structure came from Deep Convolutional Generative Adversarial Network (DCGAN).[9] The code snip written by Erik Linder-Norén can be find in https://github.com/eriklindernoren/PyTorch-GAN/tree/master/implementations/dcgan. Please refer to the GitHub repository of FLGAN.

3.2 Dataset

We will evaluate our model on three datasets: MNIST,[10] Fashion MNIST,[11] and FER-2013.[12]

The MNIST dataset contains a large number of handwritten digits. Machine learning researchers love the MNIST dataset and use it as a benchmark to validate their algorithms. For almost all algorithms, MNIST is the first dataset researchers would try; however, we wanted to extend the model use to

6. To see code for the model, please check out the repository https://github.com/yangzi33/flgan.
7. Goodfellow et al., "Generative adversarial nets."
8. Radford, Metz, and Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks."
9. Radford, Metz, and Chintala.
10. Deng, "The mnist database of handwritten digit images for machine learning research."
11. Xiao, Rasul, and Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms."
12. Goodfellow et al., "Challenges in representation learning: A report on three machine learning contests."

more datasets, since MNIST is too simple.[13] Convolutional nets can achieve 99.7% accuracy on classification tasks using MNIST, and many algorithms can perform very well easily (above 97%). Fashion MNIST is a great alternative dataset for an initial model benchmark: it contains a large number of grey-scaled images, labeled with 10 classes of clothing categories. Additionally, we used the FER-2013 dataset as an extended benchmark. The FER-2013 dataset contains images of faces labeled with 7 categories of emotions, which has more complex features.

3.3 Loss Functions

In the paper "Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks", Goodfellow[14] used minmax loss to train the DCGAN model.

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \tag{1}$$

where the generator (G) tries to maximize the probability that the discriminator D makes a mistake and the discriminator tries to minimize the probability that the discriminator itself makes a mistake.

Notice that the minimax loss is derived from the cross entropy loss:

$$CE(p,y) = \begin{cases} -\log(p) & \text{if } y = 1 \\ -\log(1-p) & \text{otherwise} \end{cases} \tag{2}$$

In reality however, it can be challenging to train datasets with various features, and cross entropy loss does not pay attention to such cases. Focal loss proposed by Tsung-Yi Lin[15] can be a solution for such problem. The focal loss puts more focus on the features that are hard to train by adding a factor $-(1-p_t)^\gamma$, where γ is a hyperparameter:

$$FL(p_t, y) = -(1-p_t)^\gamma \log(p_t) \tag{3}$$

where

$$p_t = \begin{cases} p & \text{if } y = 1 \\ 1-p & \text{otherwise} \end{cases} \tag{4}$$

Thus, we proposed to use the focal loss and modify the minimax loss to create a minimax focal loss the same way as Goodfellow derive from the cross entropy loss in GAN.[16] After modification, we get the following minimax Focal Loss:

$$\begin{aligned} \min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}}&[(1-D(x))^\gamma \log D(x)] \\ &+\mathbb{E}_{z\sim p_z(z)}[D(G(z))^\gamma \log(1-D(G(z)))] \end{aligned} \tag{5}$$

Notice that for minimax loss, $log(1 - D(G(z)))$ saturates and that the model can get stuck. Thus, for DCGAN, Goodfellow trains the generator to maximize $logD(G(z))$. The model is as follows: for the generator, maximize

$$\log D(G(z)) \tag{6}$$

for the discriminator, maximize

$$\log D(x) + \log(1 - D(G(z))) \tag{7}$$

Notice that Focal loss is a modified version of Cross entropy loss, thus, $D(G(z))^\gamma \log(1 - D(G(z)))$ should also saturate and causes stucking. Thus for FLGAN, we choose to maximize $(1 - D(G(z)))^\gamma \log D(G(z))$.

For our generator, we maximize

$$(1 - D(G(z)))^\gamma \log D(G(z)). \tag{8}$$

13. Xiao, Rasul, and Vollgraf, "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms."
14. Goodfellow et al., "Generative adversarial nets."
15. Tsung-Yi Lin et al., "Focal loss for dense object detection," in *Proceedings of the IEEE international conference on computer vision* (2017), 2980–2988.
16. Goodfellow et al., "Generative adversarial nets."

For the discriminator, we maximize

$$(1 - D(G(x)))^{\gamma} \log D(x) + D(G(z))^{\gamma} \log(1 - D(G(z))) \tag{9}$$

Also note that we have used $\gamma = 2$ suggested by Tsung-Yi Lin et al. and we choose to not use a α-balanced variant.

All the optimizations were done iteratively using Adam.[17]

3.4 Algorithm

Our algorithm for FLGAN:

Algorithm 1 Training of focal loss generative adversarial nets. In our own experiment, we set the batch size to be $batch\_size = 64$, the learning rate to be $\alpha = 0.0002$, the dimension of latent space $latent\_dim = 1000$, and $\gamma = 2$

. We sample from the generator every 1000 iterations.

$dataloader \leftarrow$ shuffled iterations of groups of batch_size images
for batch_size **do**
  for images in $dataloader$ **do**
    $m \leftarrow batch\_size \times latent\_dim$
    Draw m **training examples** $\{x^{(1)}, x^{(2)}, \ldots, x^{(m)}\}$ **from the data distribution** p_{data}
    Draw m **random noise samples** $\{z^{(1)}, z^{(2)}, \ldots, z^{(m)}\}$ **from the noise distribution** p_z**.**
    Update the generator by minimizing generator loss using Adam optimizer with learning rate α:

$$L^{(G)} = -\sum_{n=1}^{m} (1 - D(G(z^{(n)})))^{\gamma} \log D(G(z^{n})) \tag{10}$$

    Update the discriminator by minimizing discriminator loss using Adam optimizer with learning rate α:

$$L^{(D)} = -\sum_{n=1}^{m} [(1 - D(G(x^{n})))^{\gamma} \log D(x^{n}) + D(G(z^{n}))^{\gamma} \log(1 - D(G(z^{n})))] \tag{11}$$

  end for
end for

---

4 Experiments: Result and Evaluation

All experiments were implemented using Python and run on a Macbook workstation and was evaluated on three different datasets – MNIST, Fashion MNIST, and FER-2013 – as mentioned earlier in Section 3.2.

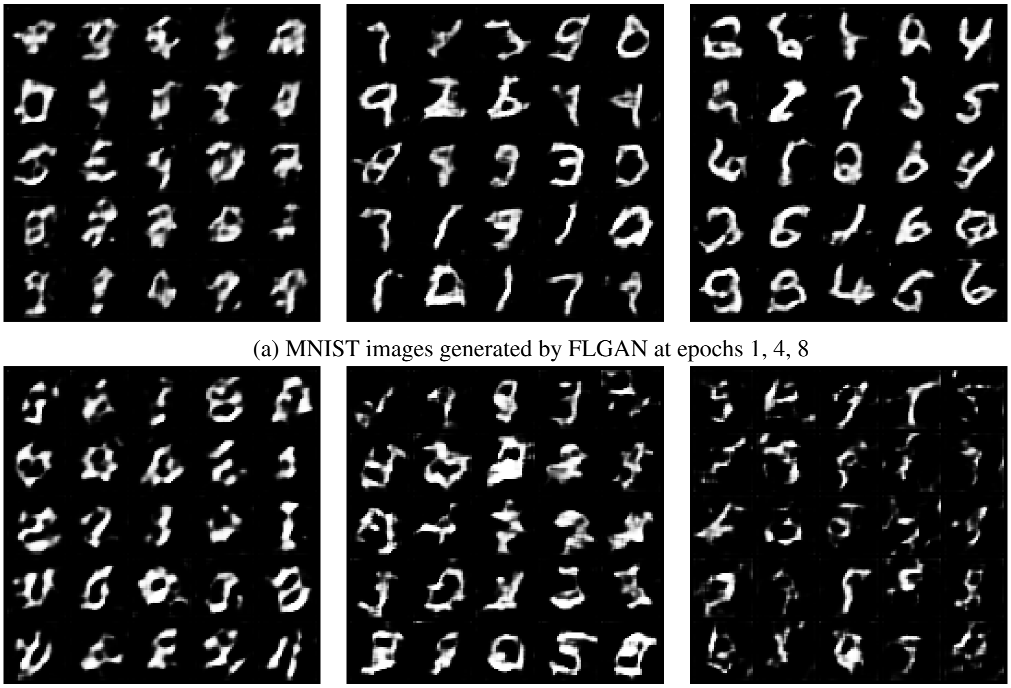
4.1 MNIST

As shown in figure 2, images produced by FLGAN is better than those produced by DCGAN at epochs 1,4,8 respectively. We can distinguish numbers like 5, 4, 6, 9 easily at epoch eight for FLGAN but we can barely see any symbol in epoch eight generated by DCGAN. Meanwhile, we can see that FLGAN is approaching convergence a lot faster than DSCAN, which proves that FLGAN can efficiently cut convergence rate. We observed that it (FLGAN) took about half of the original time to converge compared to DCGAN.

4.2 Fashion MNIST

Similar results are obtained for the dataset Fashion MNIST. As we can see from figure 3, at epoch fifteen, there are only two clothes for which we cannot distinguish its category produced by FLGAN

17. Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014,



(a) MNIST images generated by FLGAN at epochs 1, 4, 8

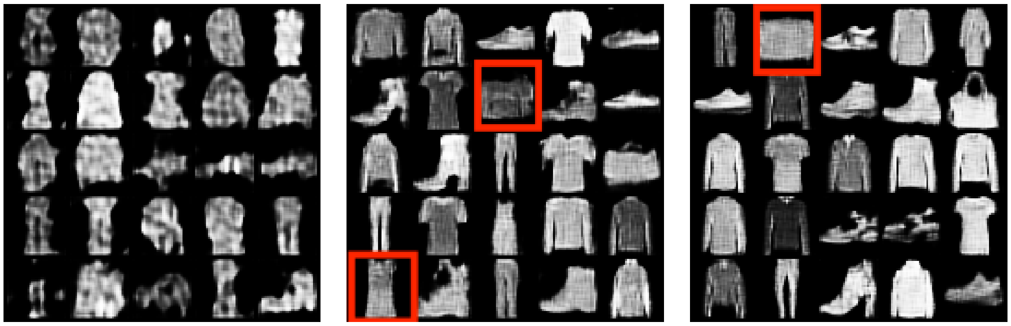
(b) MNIST images generated by DCGAN at epochs 1, 4, 8

Figure 2: Image generated using MNIST

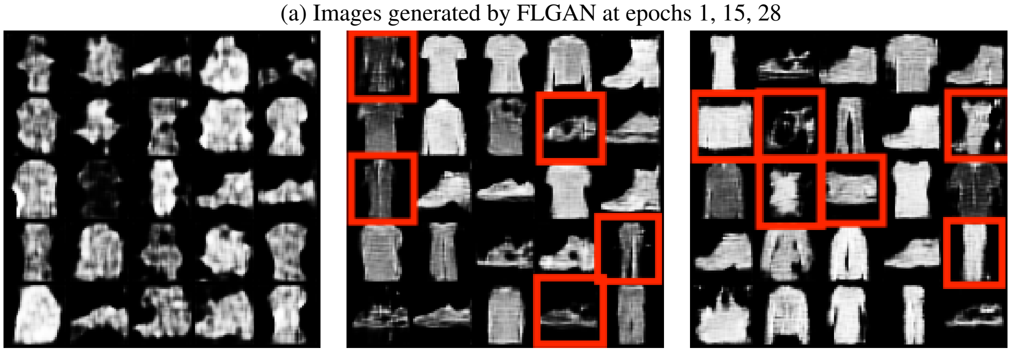
(indicated in the red square on the image) whereas there are a total of five indistinguishable images produced by DCGAN. As the number of iterations increases, this comparison gets even clearer to be observed. At epoch 28, only one item category is not observable but there are a total of 6 DCCGAN produced images that cannot be identified with their category. This clearly shows that FLGAN does produce better results and converges faster as each of its results at each epoch is always better than the image produced by DCGAN at that given epoch.

4.3 FER-2013

Results from FER-2013 are shown in figures 4 and 5. As we can see, image quality produced by FLGAN is much better than whats of DCGAN, taking epoch 400 as a example. We can recognise most images as faces for FLGAN's result but can only recognize a few in DCGAN's production, which makes results of FLGAN relatively more realistic, hinting a better model performance. Also, it seems like it is about to converge at epoch 400 for FLGAN but not for DCFGAN, implying a shorter convergence time for FLGAN.



(a) Images generated by FLGAN at epochs 1, 15, 28



(b) The generated images of DCGAN at epochs 1, 15, 28

Figure 3: Image generated using Fashion MNIST

Figure 4: Generated images of DCGAN at epochs 1, 15, 28, 200, 400

Figure 5: Generated images of FLGAN at epochs 1, 15, 28, 200, 400

5 Conclusion and Future Work

In conclusion, all experiments have shown that our proposed FLGAN model does greatly improve the image quality while speeding up the convergence time needed. There are no evident vanish gradients or mode collapse happening, which are both common problems with GAN and DCGAN models. This further shows the effectiveness of FLGAN compared to DCGAN and GAN.

However, notice that there are limitations to our proposed FLGAN model. One problem we encountered during the training process is that although result quality is overall increasing as the number of iterations increases, the quality of results is quite unstable – sometimes noisy results are generated after gaining a high-quality result several iterations earlier. Future research is suggested to work on resolving this problem. In addition, as one might notice, our model does not work as well as we expected on faces. One potential reason is that there is no labelling in GAN/DCGAN/FLGAN. Thus, it would require more layers to learn faces than what we chose to do in the experiment. It is suggested for future research to add more layers to the model or make other changes to make the model better at generating realistic faces.

6 Author Contributions:

- Conceptualization
  - Proposing ideas: Ziyue Yang, Anny Day, and Ruyi Qu.
  - Researching and proposing the focal loss function: Ruyi Qu and Anny Dai.
- Researching Related Work: Anny Dai, Ziyue Yang, and Ruyi Qu
- Experimenting and Drafting: Anny Dai, Ruyi Qu and Ziyue Yang
- Training models using a dedicated GPU: Ziyue Yang.
- Formatting and Editing in the LaTeXenvironment: Ziyue Yang.

References

Deng, Li. "The mnist database of handwritten digit images for machine learning research." *IEEE Signal Processing Magazine* 29, no. 6 (2012): 141–142.

Goodfellow, Ian, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets." *Advances in neural information processing systems* 27 (2014).

Goodfellow, Ian J, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, et al. "Challenges in representation learning: A report on three machine learning contests." In *International conference on neural information processing*, 117–124. Springer, 2013.

Kingma, Diederik P, and Jimmy Ba. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980*, 2014.

Lin, Tsung-Yi, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. "Focal loss for dense object detection." In *Proceedings of the IEEE international conference on computer vision*, 2980–2988. 2017.

Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." *arXiv preprint arXiv:1511.06434*, 2015.

Xiao, Han, Kashif Rasul, and Roland Vollgraf. "Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms." *arXiv preprint arXiv:1708.07747*, 2017.