Reproducibility in Machine Learning

Moyang Chen [260787220], Nianzhen Gu [261044523], Yiyuan Shang[260833831]

December 13, 2021

1 Introduction

In this paper, we do the reproducibility work on the paper GRADIENT ORIGIN NETWORKS [1], which is published as a conference paper at ICLR 2021.

First, we introduce the scope of our reproducibility, which includes two claims from the original paper. Second, we describe the methods that we used to reproduce the claims. Then, we compare our results to the original paper to see whether we can get the same conclusions. Finally, we discuss the easy and difficult parts of our approach.

2 Scope of reproducibility

In this paper, the authors propose a new type of generative model called Gradient Original Networks (GON) that can quickly learn a latent representation without an encoder. It is achieved by introducing a zero vector z as the latent variable and using one gradient descent step to update its value. It shows that the GON converges fast, with significantly low reconstruction error, while requiring a small number of parameters. There are two claims in the paper that we find interesting:

- claim1: GON achieves the lowest reconstruction loss on MNIST, Fashion-MNIST, Small NORB, and CIFAR-10 datasets when compared to other models.
- claim2: The ELU non-linearity activate function is more effective than other activate functions in the GON network.

3 Methodology

We used authors' code for training and write our test code to reproduce the results regarding our claims above. Because there is only training code available in GitHub.

3.1 Model descriptions

In the experiment, we chose GON as our main model. This model consists only of a decoding network, which involves 4 convolution layers with Batch Normalization, and the ELU non-linearity as activation function. z is approximated using empirical Bayes in a single step.

3.2 Datasets

There are 6 datasets in the original paper. However, we do experiments on 2 datasets, including MNIST [2] and Fashion-MNIST [3]

The MNIST dataset is of small square 28×28 pixel grayscale images of handwritten single digits between 0 and 9. The MNIST database contains 50,000 training images, 10,000 validation images, and 10,000 testing images.

Fashion-MNIST is meant as a drop-in replacement for MNIST and is based on clothing items pictures by Zalando.

Both datasets have the same size, image structure, and 10 classes(the ten digits 0 to 9 for MNIST, and ten kinds of clothing items for Fashion-MNIST). Images are resized to 32x32 before training.

3.3 Experimental setup and code

For the first claim, we aim to get the same loss value as proposed in the paper. Because of the long training time, we cannot compare GON to other models. After training GON, we run the test code written by ourselves. Because the original paper does not mention the variables explicitly, we change the dimension of z to train the GON model.

For the second claim, comparisons are done between different activation functions when training GON model. We used the metric of reconstruction loss along with training epoch number to compare the performance of activation functions, including ELU, ReLU, tanh, LeackyReLU and Softplus.

Finally, we make an improvement by increasing the gradient descent times to improve the GON model.

3.4 Computational requirements

The experiments were run on a single GPU. In the GON model, it took about 2 hours to train MNIST and Fashion-MNIST dataset. The larger the latent representation, the longer training time it takes.

4 Results

4.1 Result 1

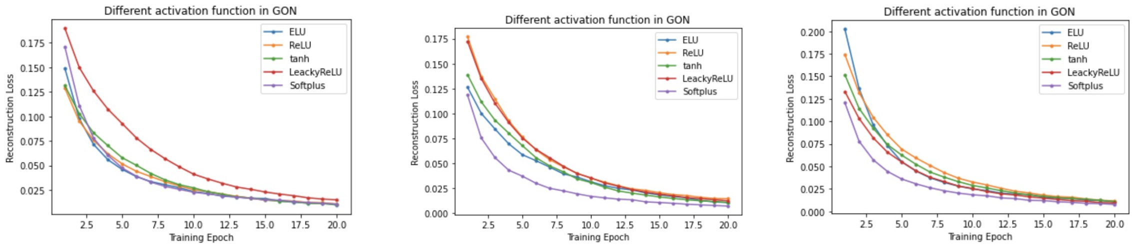
Table 1 shows our results. The representation becomes better and better if we increase the size of z. We can get similar results (the loss value when z is 256) as the original paper. However, we are not sure whether we are using the same parameter values.

Validation Reconstruction Loss		
Latent representation dimensions	MNIST	Fashion-MNIST
64	0.59	2.40
128	0.51	1.36
256	0.40	1.10

Table 1: The validation reconstruction loss by applying different dimension of latent variable (z) after 500 epoch training.

4.2 Result 2

In the original paper, the authors use ELU, which gives them the average lowest training loss. We use the same way to train the GON model and select the best activation function. Figure 1 shows the result of applying different activation functions to train the MNIST dataset. Generally speaking, ELU can give us a better result even though our result is a bit different from the original paper in terms of the shape of the curve.



(a) 32 latent variables (b) 128 latent variables (c) 512 latent variables

Figure 1: The impact of activation function and number of latent variables (z) on model performance for a GON measured by comparing reconstruction losses through training.

4.3 Results improved

In the original paper, the author only do one gradient descent to update z then use it to train the GON model. We believe that z could be a better latent representation if we do more gradient descents to update its value. Therefore, we update z twice and train the GON model. Figure 2 shows that our proposal can achieve a lower training loss.

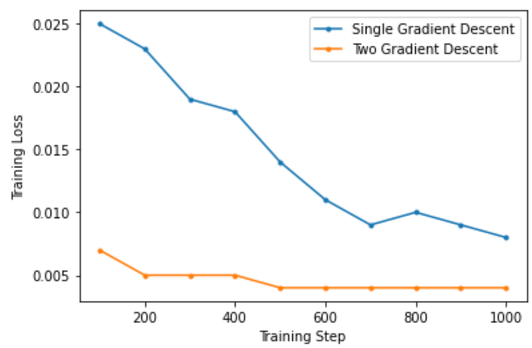


Figure 2: The training loss comparison between single and two gradient descent to latent representation. Loss is calculated in one batch (64 data).

5 Discussion

5.1 Easy and difficult parts

The code provided by the authors is easy to follow because they are arranged straightforwardly way. Also, the loading dataset step is easy since MNIST and Fashion-MNIST can be downloaded directly using code. Therefore, the second claim is easy to verify.

However, as we mentioned above, only the training code is provided in GitHub. The difficult thing is we need to write the testing code to verify the second claim. Therefore, we need to try different combinations of hyperparameters, which is time-consuming.

5.2 Communication with original authors

The corresponding code of the original paper is limited in GitHub. Lots of results cannot be reproduced using these codes. Therefore, we asked questions on GitHub. Also, we sent an email to the first author of this paper. However, we didn't get any reply.

References

[1] Sam Bond-Taylor and Chris G. Willcocks. Gradient origin networks. In *International Conference on Learning Representations*, 2021.

[2] Yann LeCun, Corinna Cortes, and CJ Burges. Mnist handwritten digit database. *ATT Labs [Online]. Available: http://yann.lecun.com/exdb/mnist*, 2, 2010.

[3] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *CoRR*, abs/1708.07747, 2017.