Theoretical Study of Distributed Co-evolutionary WGANs

Kaitav Mehta
Intro to Artificial Intelligence (Dr. Ziad Kobti)
University of Windsor
Windsor, Canada
mehta128@uwindsor.ca

***Abstract*—Recent developments in Deep Learning domain is mind boggling when it comes to learning probability distribution in neural nets, and one of key part for such progress is because of Generative Adversarial Nets(GANs). In which two neural network (Generator and Discriminator) compete among each other to learn probability distribution of points in visual pictures. Lots of research has been done to overcome the shortage of GANs like training instability, mode collapse and convergence of neural nets. One powerful technique to overcome these shortcoming is Wasserstein- GAN which become dominant method for deep generative modelling, even though in WGANs the training stability improves but it can have non convergent limit cycle near equilibrium. A new System called Lipizzaner was introduced which uses power of gradient-based GANs and as a supplement coevolution process, a class of black-box (gradient free) co-optimization technique. Experiments in Lipizzaner states that coevolution is useful framework for dodging GAN training behaviour. Even though synthesizing natural images (ImageNets) remains a next goal to be achieved using evolutionary techniques. This study shows a technique which can be useful for achieving this goal using coevolution and WGANs.**

***Index Terms*—coevolution, WGAN, ImageNet, Generative Adversarial Network , evolutionary GAN**

I. Introduction

Generative Models aims to learn functions that express distributional outputs. In normal process, generative models takes training dataset and try learn to represent an estimate of that distribution. Because of the estimate distribution we are able to achieve different variations of images which were used as dataset images. For example if human faces were used to train GANs [1] they will try to reproduce mixture of images with achieving different human faces. In recent developments of GANs they have been used on various image datasets like numbers, human faces, geographic location images etc. In normal conditions of GAN there will be one Generator Artificial Neural Network (NN) and Discriminator Artificial Neural Network. The generator will aim to learn estimate probability distribution of a given data and discriminator NN will try to identify if the generated image is real or not. Despite success of GANs it is well known that GANs are difficult to optimize. Also in Game Theory perspective the generators and discriminator are compared with two-player minimax game. In which both generators and discriminators try to achieve their respective goals, it is shown that equilibrium is necessary condition in two player game [10] but in practice gradient-based GAN training often oscillates without ultimately reaching an equilibrium. Also there are certain problems with the GANs which were explored by various researchers like mode collapse, discriminator collapse, vanishing gradient [9].

To overcome these problems Martin Arjovsky devised a novel approach called as Wasserstein Generative Adversarial Networks [2], In which they altered the traditional technique of GANs and were able to improve the stability of learning, get rid of problems like mode collapse and provide meaningful learning curves. In WGANs they have explained how the earth mover's equation is used to compute the distance between two distribution. But there is also drawback to WGANs, in which it can have non-convergent limit cycles near equilibrium [11].

Two player minimax blackbox optimization and games have been a deeply researched topic for upgrading evolutionary process [12]. It is found that sorting programs can be more efficient when developed with competitive coevolving method versus other evolutionary approaches [13]. Likewise, Herrmann proposed a two space genetic algorithm as a general technique to solve minimax optimization problems [14]. Once can see that the Nash Equilibrium solution concpet in coevolutionary literature [8] is not that different from the approach of GAN mixtures. [10]

In this paper I propose a upgraded technique of distributed coevolutionary algorithm proposed by Abdullah called as Lipizzaner [3], with advance WGANs, asking weather this approach can be used to generate high quality images when tested with ImageNet datasets. The motivation behind this work is of two fold. Firstly I wanted to add my contribution in the advancements of GANs, learn the basics of research techniques and get the experience of jumping into highly researched field and improvising the performance of it. And secondly there is ample amount of research done in evolutionary techniques to solve the gradient based problems like focusing, relativism and loss of gradients [15] [16], so the goal of this paper is to add a contribution to bridge gap between evolutionary computing and deep learning towards a better understanding of gradient based and gradient free WGANs dynamics.

I report the following contribution: WGAN Lipizzaner, a coevolutionary framework to train WGANs with Earth

Mover's distance based mutations for neural net parameters and understand mixture of WGANs.

II. Related Work

A. Training GANs with evolutionary framework.

Several evolutionary gradient-based GAN training variants have been proposed. One variant category is focused on improving training technique for generator population using variation, evolution and selection technique while keeping just one vanilla discriminator like in original GAN. In which generator population is evolved using different mutation functions like minimax mutation, Heuristic mutation and Least-squares [4]. They also stated under minimal assumption that original Wasserestein GAN [2] and its variant [17] methods convincingly reduce the training instability because of wasserestein distance being continuous and differential everywhere.

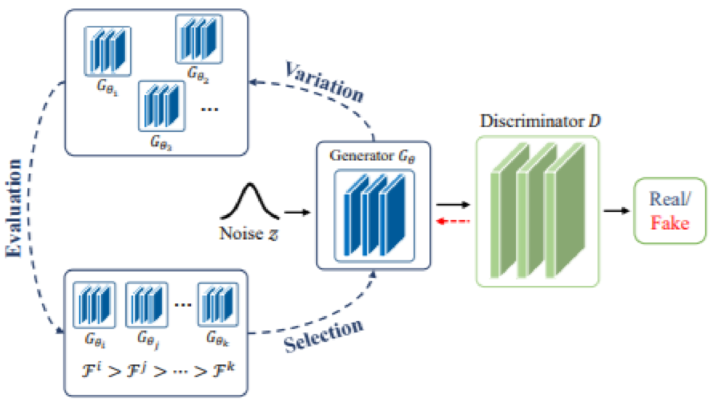

Fig. 1. Evolutionary-GAN [4]

Another approach called as Lipizzaner [3] was proposed and had notable success. They have used a spatial grid framework for coevolving both generator and discriminator. In a which asynchronous evolving of population was done based on spatial location in a grid. They also successfully dodged mode collapse when compared with DCGAN. At the end there will be a group of Generators and Discriminator which will be outcome of this approach.

We can see that all the proposed evolutionary approaches revolves around vanilla GANs and advance version of this WGANs is never been tested before to improve the quality of the generated probability distribution.

B. Wasserstein GANs

[2] argues that the divergences which GANs typically minimize are potentially not continuous with respect to the generators parameters, leading to training difficulty. They propose instead using the Earth-Mover (also called Wasserstein-1) distance W(q, p), which is informally defined as the minimum cost of transporting mass in order to transform the distribution q into the distribution p (where the cost is mass times transport distance). Under mild assumptions, W(q, p) is continuous everywhere and differentiable almost everywhere. The WGAN

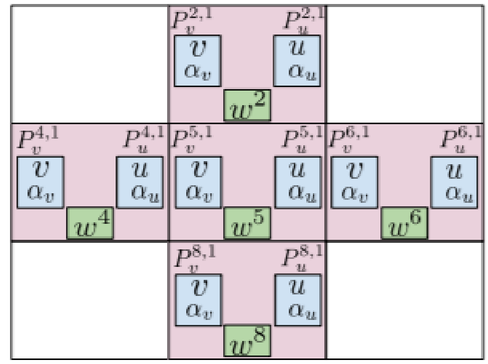

Fig. 2. Topology of a 3 3-grid (m = 3) with a neighborhood size of sn = 5. A neighborhood of the 5th cell is highlighted in light red. Each cell has a population size of one (one generator Gu and one discriminator Dv). The corresponding neural net parameters u and v are updated with gradient-based mutations, while the respective hyperparameters (e.g., learning rate u and v) are updated with Gaussian-based mutations based on the interactions of each cell with its neighbors. Each cell has the mixture weight vector wk for their respective neighborhood, which is optimized with an evolutionary algorithm according to a given performance metric [3]

value function is constructed using the Kantorovich-Rubinstein duality to obtain

$$\min_G \max_{D \in \mathcal{D}} \mathop{\mathbb{E}}_{\boldsymbol{x} \sim \mathbb{P}_r} \big[D(\boldsymbol{x})\big] - \mathop{\mathbb{E}}_{\tilde{\boldsymbol{x}} \sim \mathbb{P}_g} \big[D(\tilde{\boldsymbol{x}}))\big]$$

Fig. 3. Wasserstein equation [4]

where D is the set of 1-Lipschitz functions and Pg is once again the model distribution implicitly defined by x˜ = G(z), z ˜ p(z). In that case, under an optimal discriminator (called a critic in the paper, since its not trained to classify), minimizing the value function with respect to the generator parameters minimizes W(Pr, Pg). The WGAN value function results in a critic function whose gradient with respect to its input is better behaved than its GAN counterpart, making optimization of the generator easier. Additionally,WGAN has the desirable property that its value function correlates with sample quality, which is not the case for GANs. To enforce the Lipschitz constraint on the critic, [2] propose to clip the weights of the critic to lie within a compact space [c, c]. The set of functions satisfying this constraint is a subset of the k-Lipschitz functions for some k which depends on c and the critic architecture. [17]

C. Coevolutionary algorithms for Minimax problems.

Various coevolutionary techniques have been proposed to solve minimax problem in different domains. It was found that asymmetry can be the result of GAN mode collapse [1]. To address this issue, asymmetric fitness evaluation was presented in [12]. Also attempts have been made to overcome the limitations of existing coevolutionary approach in solving minimax optimization problems using differntail evolution [18].

III. APPROACH

My approach here is to upgrade Lipizzaner framework with better adversarial nets called as WGANS. In traditional structure Gaussian based mutations are done to update generators which we will update by doing random mutations like Minimax mutation, Heuristic mutation and Lease-squares mutation as described in [4] and to train the generators and critics we will use differentiating Wasserstein distance $W(\mathbb{P}_r,\mathbb{R}_\theta)$ by back propagating the Wasserstein equation.

Data: P_u: generator population
P_v: critic population
α_i: selection probability
w: critic parameters
m: side length of spatial square grid
I: number of population generations per training step
Result: $P_u{}^*$: evolved generator mixture
$P_v{}^*$: evolved critic mixture

repeat
/*spatial Coevolution of Generator and Critic Population*/
while *parfor k in range(m^2)* **do**
$\hat{P}_u{}^k,\hat{P}_v{}^k \leftarrow BasicCoevWGANs(P_u{}^k, P_v{}^k, w,\alpha_i,I)$
check Lipizzaner for Basic Coevolve Algo [3]
Find the fitness of each generator based on fitness function of diversification and optimal discriminator loss function as stated in EGAN to get the best generator and critic.
$P_u{}^k \leftarrow TopN(\hat{P}_u{}^k)$
$P_v{}^k \leftarrow TopN(\hat{P}_v{}^k)$
Mutate Generators which are not fit with minimax, heuristic or lease-squares gradient based mutation//
end
Until training is converged

Algorithm 1: Coevolving Distributed WGANs

In this algorithm individuals from both the populations are distributed spatially, with local interactions governing fitness evaluation, selection and mutation. As shown in original paper, spatial coevolution to be substantially successfully over several non trivial learning task due to it's ability to maintain diversity over long run. Here notable difference is that only the centre of the spatial grid is updated on each run instead of updating all the population of the selected grid. And selection grid will keep moving after every update. Since there are m^2 neighbourhoods, all the individual population will get updated.

After the training of generator and critic is converged, a particular spatial grid as a whole is evaluated and the best one is used for generating probability distribution over given dataset. Because of the diversity of it's arrangement, training stability achieved because of gradient based wasserstein distance back propagation and E-GAN mutation technique I assume that if this algorithm is properly implemented than it can achieve better performance results over traditional Lipizzaner.

IV. EXPERIMENTAL SETUP

As the title states this is just theoretical study, as the experimental procedure for creating lipizzaner is complicated and not all details are clearly mentioned in the original paper.

With that stated even though I have carried out baby steps for creating basic version of coevolving GANs. Before jumping directly into coevolving GANS, I have carried out simple experiments of predator prey coevolution and basic models of vanilla GANs in python.

In my experiment of coevolving GANs, there there will be 3 count of Generators and Discriminators and both the population of generator and discriminator will try to coevolve by using gradient based mutations after every generation. I have carried out this experiment using pytorch framework in Python and performed experiment in MNIST standard dataset which is an open source dataset. The experiment will be running for 100 generations and each generation will have 16 sample data. The experiment took 3-4 hours to process all the generations and give the final results. I will attach my code with the paper so if one wants to carry out experiment then it can be possible.

In the experiments results I have found that the generators were learning at faster rate as compared with vanilla GANS, the figure shows the mode collapse in my model after running 1000 generations and the results of Lipizzaner framework stated in the original paper. Moreover, I have found that few population of Generator collapsing after initial few generations.

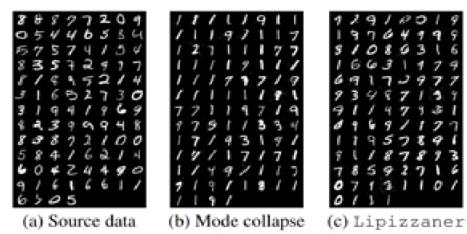

Fig. 4. My results vs Lipizzaner [3]

V. CONCLUSION

From the theoretical study and GANs literature review we can see that WGANs have better training stability and from experimental results of basic coevolving GANs and Lipizzaner framework it is evident that the competitive coevolving GANs shows improvements in mode collapsing problem. Also with the additional variation of mutating GANs as stated in EGAN(better diversification) and combination of Lipizzaner(better mode collapse) with WGAN(better training stability), I confidently assume that there is a high chance for better performance in generators when tested on ImageNet and can get better Inception score.

VI. FUTURE WORK

For the future work I successfully want to implement the stated theoretical model of WGAN Lipizzaner and compare results with the original WGAN and Lipizzaner models. If the results are found to be significant then I am planning to publish paper in good conference like GECCO 2019 or Canadain AI 2019.

ACKNOWLEDGMENT

I highly acknowledge Dr Kobti for motivating and giving right research direction to me through out this research project.

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 (pp. 2672-2680).

[2] Arjovsky M, Chintala S, Bottou L. Wasserstein generative adversarial networks. In International Conference on Machine Learning 2017 Jul 17 (pp. 214-223).

[3] Schmiedlechner T, Al-Dujaili A, Hemberg E, O'Reilly UM. Towards distributed coevolutionary gans. arXiv preprint arXiv:1807.08194. 2018 Jul 21.

[4] Wang C, Xu C, Yao X, Tao D. Evolutionary Generative Adversarial Networks. arXiv preprint arXiv:1803.00657. 2018 Mar 1.

[5] Brock A, Donahue J, Simonyan K. Large scale gan training for high fidelity natural image synthesis. arXiv preprint arXiv:1809.11096. 2018 Sep 28.

[6] Hillis WD. Co-evolving parasites improve simulated evolution as an optimization procedure. Physica D: Nonlinear Phenomena. 1990 Jun 1;42(1-3):228-34.

[7] Mitchell M. Coevolutionary learning with spatially distributed populations. Computational intelligence: principles and practice. 2006.

[8] M. A. a. K. A. D. J. Potter, "A cooperative coevolutionary approach to function optimization," 1994.

[9] Arora S, Zhang Y. Do GANs actually learn the distribution? an empirical study. arXiv preprint arXiv:1706.08224. 2017 Jun 26.

[10] Arora S, Ge R, Liang Y, Ma T, Zhang Y. Generalization and equilibrium in generative adversarial nets (gans). arXiv preprint arXiv:1703.00573. 2017 Mar 2.

[11] Nagarajan V, Kolter JZ. Gradient descent GAN optimization is locally stable. In Advances in Neural Information Processing Systems 2017 (pp. 5585-5595).

[12] Jensen. A new look at solving minimax problems with coevolutionary genetic algorithms. In Metaheuristics: computer decision-making, 2003.

[13] Herrmann. A genetic algorithm for minimax optimization problems. In CEC, 1999.

[14] Hillis. Co-evolving parasites improve simulated evolution as an optimization procedure. Physica D: Nonlinear Phenomena, 1990

[15] Nolfi and Floreano. Coevolving predator and prey robots: Do arms races arise in artificial evolution? Artificial life, 1998

[16] Watson and Pollack. Coevolutionary dynamics in a minimal substrate. In GECCO, 2001

[17] Gulrajani I, Ahmed F, Arjovsky M, Dumoulin V, Courville AC. Improved training of wasserstein gans. In Advances in Neural Information Processing Systems 2017 (pp. 5767-5777).

[18] Qiu et al. A new differential evolution algorithm for minimax optimization in robust design. IEEE transactions on cybernetics, 2017.