# A Study Of End-to-End Energy Based Generative Adversarial Networks

**Akarsh Pokkunuru**
Dept. of Computer Science,
University of North Carolina at Charlotte,
apokkunu@uncc.edu

**Pedram Rooshenas**
Dept. of Computer Science,
University of North Carolina at Charlotte,
rooshenas@uncc.edu

## Abstract

Energy based models (EBM) can jointly present a unified framework for probabilistic and non-probabilistic models by jointly associating their random variables to a single scalar energy value. We thus leverage this powerful property to study various Energy based Generative Adversarial Networks (GANs) and apply them to the domain of structured prediction. More specifically, we aim to understand how different types of GAN models can perform on an image segmentation task, when trained in an End-to-End discriminative setting. At the end of this project report, we present various evaluation metrics and results to showcase that our End-to-End EBM GANs have the capability to model complex structured data.

## 1 Introduction

The idea behind Energy based models [7] is simple yet, it represents a powerful machine learning philosophy. These models have the capability to jointly map a given supervised learning input space $x_i \in X$ and labels $y_i \in Y$ into an energy landscape $S$ such that, $S = E(X, Y)$ where, $E$ represents the energy model itself. We then carefully design a model and its objective function to minimize the overall $S$ using traditional gradient descent based optimization as follows:

$$\widehat{y} = arg\min_{y} E_x(y) \tag{1}$$

This methodology creates an accurate mapping of the input space to the energy space while learning a joint representation of all input variables. For this project, we aim to study the task of structured prediction such as image segmentation using a conjunctive procedure between EBMs and GANs [4], [9], [11]. Prior research [6], [12], [13] has shown great success in applying GANs in this area. This is due to the fact that, GANs can jointly represent inputs and outputs during the minimization of training objective while EBMs learn the joint representation of the input variables. Thus when these two solutions are coupled together, the resulting generator model of an energy based GAN can be capable of creating complex and realistic looking synthetic segmented images [14], [15].

Despite the aforementioned promises, GANs still fundamentally suffer from several drawbacks such as mode collapse, exploding and vanishing gradients, non-convergence between the generator and discriminator network and general training stability. In order to solve these shortcomings, many solutions were proposed in literature [16], [17]. However, unrolled optimization of GANs stands out among the rest of the solutions since this method has more possibility of finding a global minima among all the available local minima [8]. This results in a higher convergence rate and better training stability. In lieu of EBMs, unrolled optimization also offers an advantage which is, test time energy minimization. Since most NN models focus on training data inference maximization, they often lack the heuristics for test time inference leading to poor generalization. Thus given these advantages, we conduct a comparison study for various GAN models trained with an end-to-end energy based generator network and prove that the resulting model can create a smoother learning landscape and higher quality images when compared to standard GAN models.

The core contributions of this project are as follows:

- We present a comprehensive comparison study of various types of GAN models such as EB GAN, WGAN-GP, MSE GAN, Log-loss GAN for image segmentation task. In all these selected models, both the generator and discriminator networks are trained under discriminative setting.
- The discriminator network is conditioned on a combinatorial sequence of RGB images, ground truth mask and synthetic mask and thus, acts as a standard binary classifier to distinguish the predicted masks from ground truth masks.
- Similarly, the generator network is conditioned on RGB images, Gaussian noise and learns to produce synthetic masked images as predicted labels using the unrolled inference procedure. Thus, the generator is designed to act as both the unrolled inference network and the energy based model.
- We then use an exponentiated gradient descent based Langevin dynamics sampling as the inference procedure for the generator network.
- In order to strengthen the generator inference, we additionally pre-train the generator network with a Cross entropy loss at each step of the training procedure. This loss is weighted and eventually fed to the generator objective function of each type of GAN model.
- At the end of the report, we present comprehensive loss landscape visualizations and IOU metrics for analyzing and understanding the performance of each GAN model.

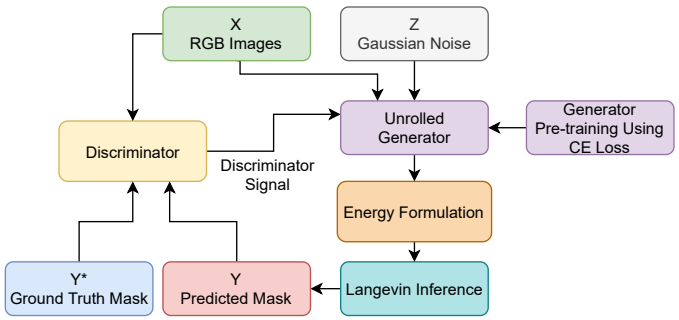
The rest of the report is organized as follows: Section 2 discusses some of the recent solutions published in literature related to our work, Section 3 and 4 discuss the project methodology and evaluation results respectively. Finally, in Section 5 we present some important discussions based on the results and duly the conclude the report.

## 2 Related work

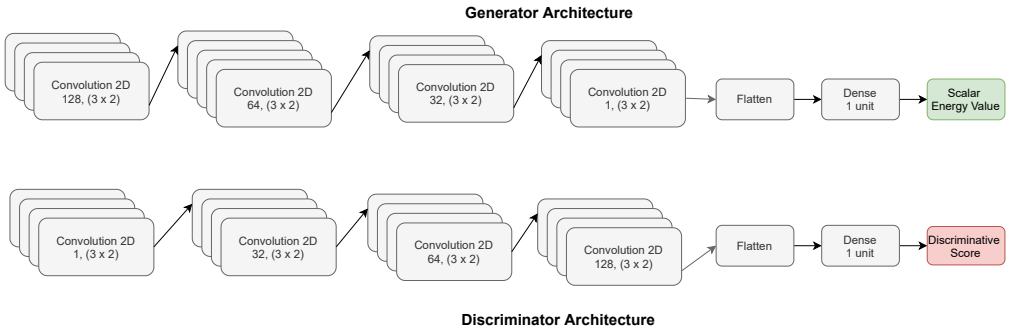
In this section, we briefly discuss some of the works available in literature related to this project. The general idea of adversarial training was introduced in [18]. Since the introduction of this model, many new developments have been proposed to solve some of the shortcomings of this original GAN. More recently, a Wasserstein distance based GAN[16] and its improved gradient penalty mechanism [17] were proposed to solve several shortcomings mentioned in the aforementioned section. Similarly, an unrolled GAN was proposed in[8] with a focus on unrolling the discriminator network with respect to the generator network objective function leading to a stable model and reduced mode collapse. From the energy based models and structure prediction task perspective, EB-GAN [6] combined the idea of energy based models and GANs in order to solve the task of image generation. The core idea behind this work was to use an autoencoder based discriminator as the energy network and assign high energy values to incorrect predictions and vice-versa for correct predictions. More importantly, SPEN [1] and End-to-End SPENs [2], [3] proposed the use of energy based models for structured prediction. These models are novel however, the authors train the energy based model using a cross entropy loss which fails to jointly model the input variables due to independence assumption of the objective function. To improve upon this shortcoming, [4],[5] were proposed to use energy based models using adversarial training. The core idea behind this work [5] is to use GANs in a energy based setting and train the network with various constraints on the input data. While, [4] designed a energy based GAN where the authors propose a generator network as the inference model and discriminator as the energy network. Our work is inspired by some of these works with some significant changes which will be discussed in detail in the coming sections.

## 3 End-to-End Energy based GANs

Our end-to-end energy based GAN framework is illustrated in Fig. 1 below. As seen in Fig. 1(a), the core idea of our model is to encompass the generator network $G$ as both the energy and inference generation procedures into one compact end-to-end NN. Initially, the discriminator network $D$ is conditioned on a combinatorial input sequence consisting of RGB images, ground truth mask and synthetic mask. It then acts as a standard binary classifier to distinguish the predicted masks from




(a) Framework overview




(b) Network Design

Figure 1: End-to-End Energy Based GAN Framework

ground truth masks. For our experiments, we chose the following standard discriminator objective functions listed below:

$$\begin{aligned}\text{WGAN-GP:} \quad \mathcal{L}_D &= -D(X, Y^*, Y^*) + D(X, Y^*, G(X, Z)) \\ &+\lambda||\nabla D(\alpha\,(X, Y^*, Y^*) + (1-\alpha\,(X, Y^*, G(X, Z)))||_2 - 1)^2\end{aligned} \tag{2}$$

$$\text{EB GAN:} \quad \mathcal{L}_D = D(X, Y^*, Y^*) - [m - D(X, Y^*, G(X, Z))]^+ \tag{3}$$

$$\begin{aligned}\text{MSE GAN:} \quad \mathcal{L}_D &= ||\alpha R(Y^*, G(X, Z)) + E(X, Y^*, G(X, Z))||^2 \\ \text{where;} \quad R(Y^*, G(X, Z)) &= \frac{\sum_i^M \min(Y_i^*, G_i(X, Z))}{\sum_i^M \max(Y_i^*, G_i(X, Z))}\end{aligned} \tag{4}$$

$$\text{Log loss GAN:} \quad \mathcal{L}_D = \log\left(1 - D(X, Y^*, G(X, Z))\right) + \log D(X, Y^*, Y^*) \tag{5}$$

here, $\lambda$, $\alpha$ in Equ. 2 represent the gradient penalty term and interpolation noise term. Additionally, $R$ in Equ. 4 represents the IOU based reward value computed between the ground truth and predicted masked labels.

In contrast to the discriminator training procedure, the generator network is conditioned only on RGB images, Gaussian noise and learns to produce synthetic masked images as predicted labels using an unrolled inference procedure. This learning procedure is to iteratively minimize the scalar energy value using an unrolled optimization procedure for every input RGB image $X$ and its corresponding synthetic mask $Y = G(X, Z)$ in the training data. This is shown as follows:

$$Y^T = Y_0 - \eta \sum_{t=1}^{T} \frac{d}{dY} E_X(Y_t) \tag{6}$$

Table 1: IOU Score Comparison

| Model | IOU Scores in % | | |
|---|---|---|---|
| | Train | Test | Validation |
| **WGAN-GP** | 98.82 | 66.45 | 62.88 |
| **Log Loss GAN** | 98.91 | 67.07 | 62.97 |
| **EB GAN** | 98.02 | 68.32 | 63.44 |
| **MSE GAN** | 95.98 | 67.71 | 63.84 |

Here, $Y_0$ is the initial noisy label prediction, the energy value $E_X(Y_i)$ is calculated using a Convolution NN as shown in the Fig. 1(b) and $\eta$ is the learning rate. This value is then minimized iteratively in Equ. 6 to generate synthetic samples $Y^T$ for $T$ steps. At each step of this procedure, the segmented output image of the previous step is used as input to the Langevin dynamics procedure which is shown below:

$$Y^{T+1} = Y^T - \frac{\delta}{2}\nabla E(Y^T, X) + \delta\mathbb{N}(0, \delta) \tag{7}$$

Note that, we use an exponentiated gradient descent approach while computing the gradients in Equ. 6 and 7. Additionally, $\delta$ is the inference rate parameter of the Langevin dynamics procedure. After this inference procedure, last predicted label at step $T$ is picked and a Sigmoid activation is applied to restrict the values between $[0, 1]$. This label is then passed to the discriminator network for updates and likewise, the generator network parameters are updated according to the standard generator objective functions for different GAN models as follows:

$$\text{WGAN-GP, MSE GAN, Log loss GAN:} \quad \mathcal{L}_G = -D(X, Y^*, G(X, Z)) + \alpha\, CE(Y^*, Y) \tag{8}$$

$$\begin{aligned} \text{EB GAN:} \quad & \mathcal{L}_G = D(X, Y^*, G(X, Z)) + \omega_{PT} PT(G(X, Z)) + \alpha\, CE(Y^*, G(X, Z)) \\ \text{where;} \quad & PT(G(X, Z)) = \frac{1}{B(B-1)} \sum_i \sum_{j \neq i} \Big( \frac{G_i(X, Z)^T G_j(X, Z)}{||G_i(X, Z)^T||\ \ ||G_j(X, Z)||} \Big)^2 \end{aligned} \tag{9}$$

In Equ. 9, $PT$ indicates the pull away loss regularization term and $\omega_{PT}$ is its corresponding weigh of this value.

## 4 Results and Evaluation

In this section, we provide performance evaluations for all the four GAN models described above. We specifically utilize these models to perform the task of image segmentation on the Weizmann horse dataset in a supervised discriminative setting. The models are then evaluated on IOU scores for training, testing and validation datasets. These scores are presented in Table 1 above. Additionally, we explore the stability capability of each models adversarial objective function by visualizing its loss landscape. We also visualize the distributions of the discriminator network's output scores for both synthetic and real samples to analyze the model's learning capability. All the GAN models are trained using a common network architecture which is a stacked convolution NN shown in Fig. 1(b). The model then uses the corresponding hyperparameters listed in Table 2 during training.

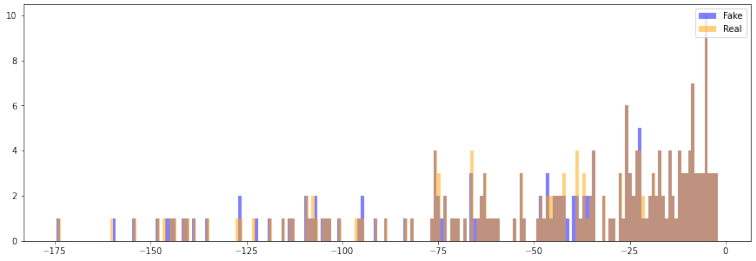
### 4.1 Unrolled Energy based WGAN-GP Model

As seen in Fig. 2(a), the discriminator scores for fake samples are indicated by blue data points and scores for ground truth samples are indicated by yellow data points. At the first epoch, the overlap between the discriminator scores for synthetic data and ground truth masks is meager. However at the end of epoch 300 as seen in Fig. 2(b), the overlap greatly improved indicating that the generator managed to learn and mimic the input data distribution. This overlapped region is indicated by the brown data points.

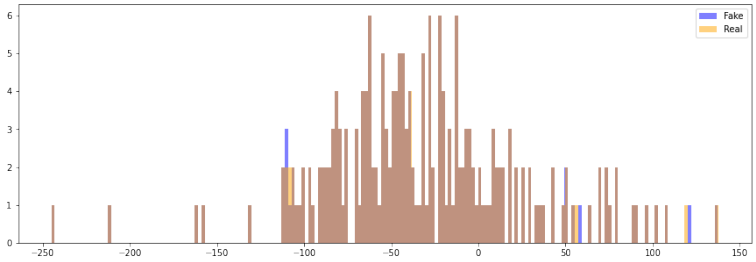
Next, in Fig. 4, we evaluate the WGAN-GP model based on visual inspection of the generated synthetic segmented masks. As seen in Fig. 4(a), the first epoch results in noisy predictions for all the three data partitions. However, at the end of the training, the generated samples are greatly improved

Table 2: Selected hyperparameters for experiments

| Parameter | Value |
|---|---|
| Generator pre-training learning rate | 0.001 |
| Generator inference learning rate | 0.0003 |
| Discriminator learning rate | 0.003 |
| Discriminator epochs | 1 |
| $B$ Batch size | 10 |
| $\delta$ Langevin inference rate | 2 |
| $T$ inference steps | 5 |
| Epochs | 300 |
| $\lambda$ WGAN gradient penalty | 10 |
| EB GAN pull-away weight | 0.01 |
| EB GAN margin | 10 |




(a) Histogram at Epoch 0




(b) Histogram at Epoch 300

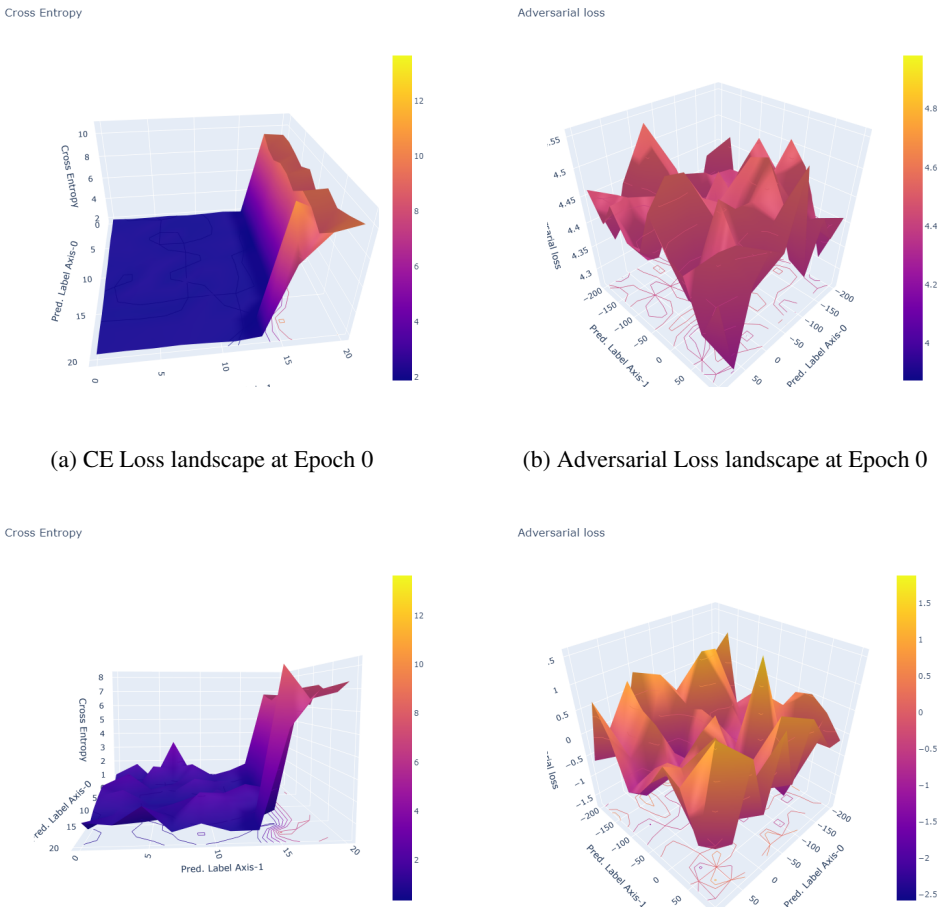
Figure 2: Histograms of Fake vs Real Discriminator scores for WGAN-GP

resulting in an improved IOU scores for train, test and validation sets as seen in Table 1. In addition, the IOU score of each epoch and all three data partitions are presented in Fig. **??**. Observing this figure, we can say that the initial fluctuations in the score values indicate the learning behavior of the generator. But after 200th epoch, the network is stabilized and produces less fluctuations indicating some form of convergence between the discriminator and generator networks.

In Fig. 3, we visualize the loss landscape for both the generator pre-training cross entropy loss function and the corresponding adversarial loss function. At the initial epoch, the fluctuations in the loss values are high however, as we approach the end of training, the landscape is smoother and indicates minimized values as seen on the loss scale.

### 4.2 Unrolled Energy based Log-Loss Model

Similar to the WGAN-GP model evaluation, we evaluate the Energy based GAN model using the metrics described above. The histogram plot for fake vs real discriminator scores is seen in Fig. 6. The corresponding synthetic samples produced by log loss GAN can be seen in Fig. 8. Finally, the loss landscapes can be viewed in Fig. 7.




(a) CE Loss landscape at Epoch 0

(b) Adversarial Loss landscape at Epoch 0

(c) CE Loss landscape at Epoch 300

(d) Adversarial Loss landscape at Epoch 300

Figure 3: Loss landscape training progression for WGAN-GP

### 4.3 Unrolled Energy based EB-GAN Model

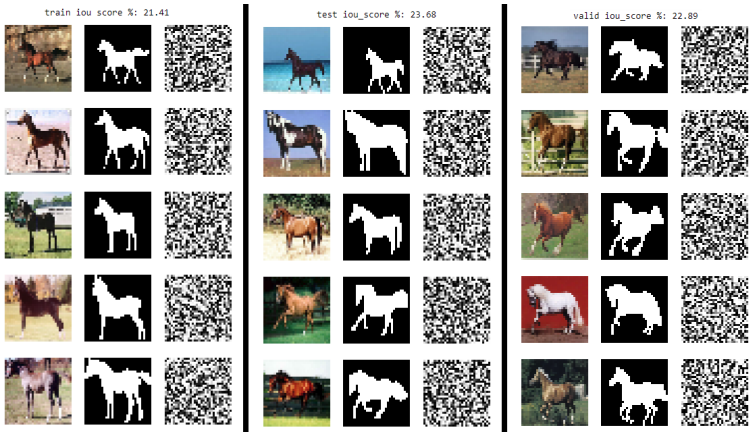
The histogram plots for fake vs real discriminator scores is seen in Fig. 9. The corresponding synthetic samples produced by log loss GAN can be seen in Fig. 11. Finally, the loss landscapes can be viewed in Fig. 10.

### 4.4 Unrolled Energy based MSE GAN Model

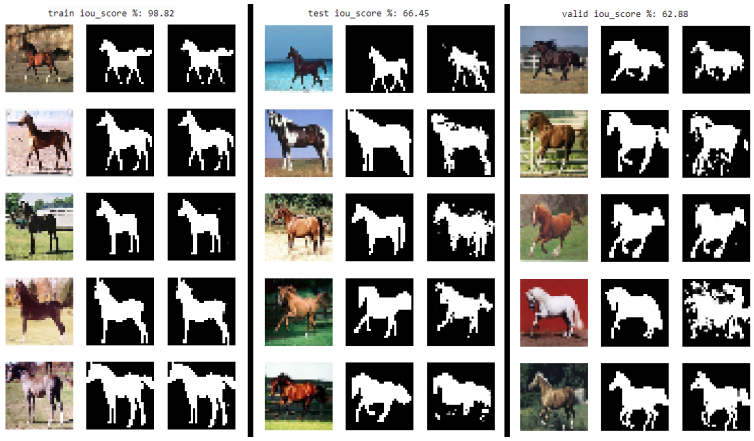
The histogram plots for fake vs real discriminator scores is seen in Fig. 12. The corresponding synthetic samples produced by log loss GAN can be seen in Fig. 14. Finally, the loss landscapes can be viewed in Fig. 13.

## 5 Discussion

Thus through this project, we explored the effect of combining an energy based model and a Langevin dynamics based inference procedure into an unrolled generator of a GAN. We then trained this end-to-end model to learn the inherent input data distributions jointly. From these results, it is clear to understand that the unrolled inference of the energy based GAN based models produce good accuracy predictions and realistic segmented images for our image segmentation task. We have to note the




(a) Samples at Epoch 0




(b) Samples at Epoch 300

Figure 4: WGAN-GP Synthetic Sample comparison for Train, Test and Valid dataset (from left to right)

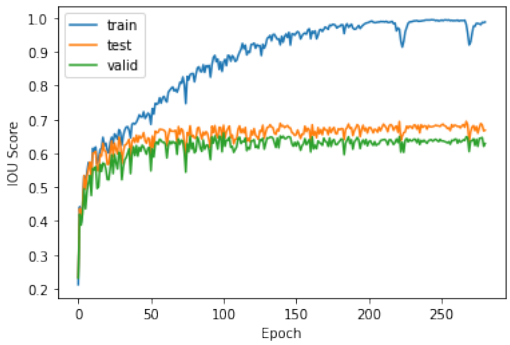
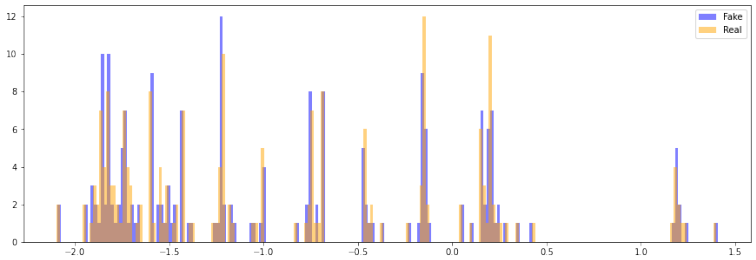
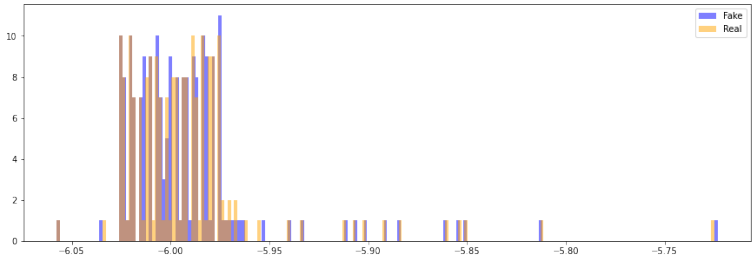



Figure 5: IOU Score progression through training for WGAN-GP




(a) Histogram at Epoch 0




(b) Histogram at Epoch 300

Figure 6: Histograms of Fake vs Real Discriminator Scores For Unrolled Energy based Log loss GAN

fact that all the GAN models use a simple and shallow representation of a convolution NN, thus seeing the true potential of the training procedure might be limited. Additionally, the Weizmann horse dataset used to conduct the experiments has limited samples and therefore, the discriminator has a high chance to overfit the data. Thus, future research can be directed towards training the model on larger datasets. Other improvements can be made towards the architecture design of the generator and discriminator. And finally, performing a randomized grid search for optimum hyper-parameters will also result in performance improvement of the predictions.

## References

[1] Belanger, D., & McCallum, A. (2016, June). Structured prediction energy networks. In International Conference on Machine Learning (pp. 983-992).

[2] Belanger, D., Yang, B., & McCallum, A. (2017). End-to-end learning for structured prediction energy networks. arXiv preprint arXiv:1703.05667.

[3] Hwang, J. J., Ke, T. W., Shi, J., & Yu, S. X. (2019). Adversarial structure matching for structured prediction tasks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4056-4065).

[4] Pan, P., Liu, P., Yan, Y., Yang, T., & Yang, Y. (2020). Adversarial Localized Energy Network for Structured Prediction. In AAAI (pp. 5347-5354).

[5] Ren, H., Stewart, R., Song, J., Kuleshov, V., & Ermon, S. (2018). Adversarial constraint learning for structured prediction. arXiv preprint arXiv:1805.10561.

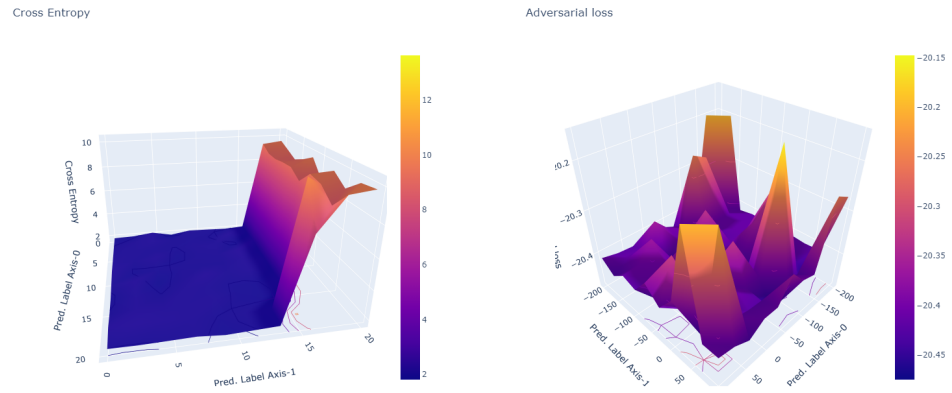
[6] Zhao, J., Mathieu, M., & LeCun, Y. (2016). Energy-based generative adversarial network. arXiv preprint arXiv:1609.03126.

[7] LeCun, Y., Chopra, S., Hadsell, R., Ranzato, M., & Huang, F. (2006). A tutorial on energy-based learning. Predicting structured data, 1(0).

[8] Metz, L., Poole, B., Pfau, D., & Sohl-Dickstein, J. (2016). Unrolled generative adversarial networks. arXiv preprint arXiv:1611.02163.

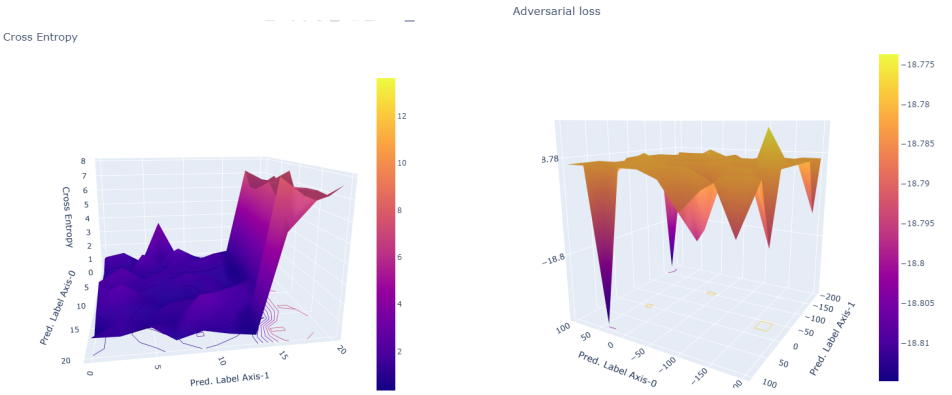
[9] Finn, C., Christiano, P., Abbeel, P., & Levine, S. (2016). A connection between generative adversarial networks, inverse reinforcement learning, and energy-based models. arXiv preprint arXiv:1611.03852.

[10] Du, Y., & Mordatch, I. (2019). Implicit generation and modeling with energy based models. Advances in Neural Information Processing Systems, 32, 3608-3618.




(a) CE Loss landscape at Epoch 0

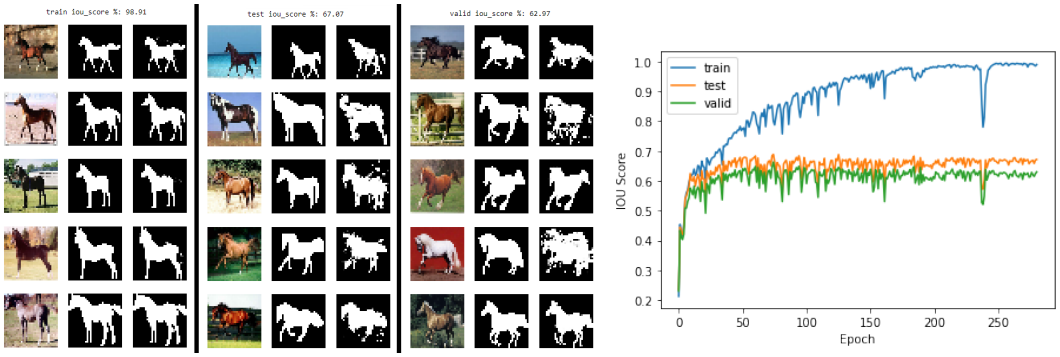
(b) Adversarial Loss landscape at Epoch 0




(c) CE Loss landscape at Epoch 300

(d) Adversarial Loss landscape at Epoch 300

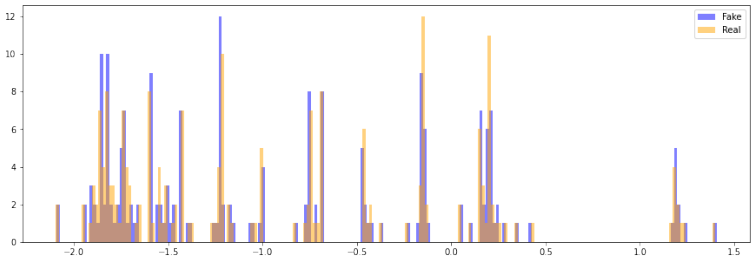
Figure 7: Loss landscape training progression for Log-loss GAN



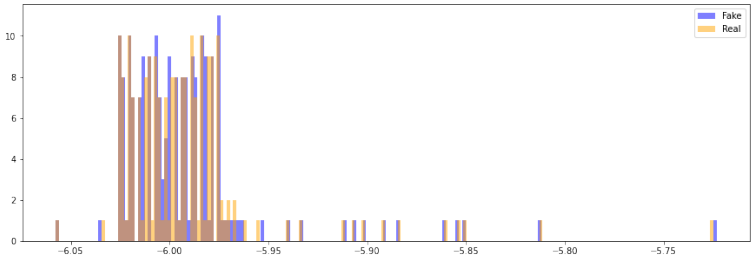

(a) Samples at Epoch 0

(b) IOU Score progression through training

Figure 8: Log-loss GAN Sample and IOU scores for Train, Test and Valid dataset (from left to right)




(a) Histogram at Epoch 0




(b) Histogram at Epoch 300

Figure 9: Histograms of Fake vs Real Discriminator Scores For Unrolled EB-GAN

[11] Dai, Z., Almahairi, A., Bachman, P., Hovy, E., & Courville, A. (2017). Calibrating energy-based generative adversarial networks. arXiv preprint arXiv:1702.01691.

[12] Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., & Terzopoulos, D. (2020). Image segmentation using deep learning: A survey. arXiv preprint arXiv:2001.05566.

[13] Luc, P., Couprie, C., Chintala, S., & Verbeek, J. (2016). Semantic segmentation using adversarial networks. arXiv preprint arXiv:1611.08408.

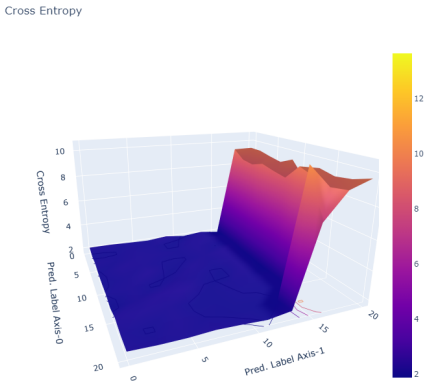
[14] Souly, N., Spampinato, C., & Shah, M. (2017). Semi supervised semantic segmentation using generative adversarial network. In Proceedings of the IEEE International Conference on Computer Vision (pp. 5688-5696).

[15] Xue, Y., Xu, T., Zhang, H., Long, L. R., & Huang, X. (2018). Segan: Adversarial network with multi-scale l 1 loss for medical image segmentation. Neuroinformatics, 16(3-4), 383-392.

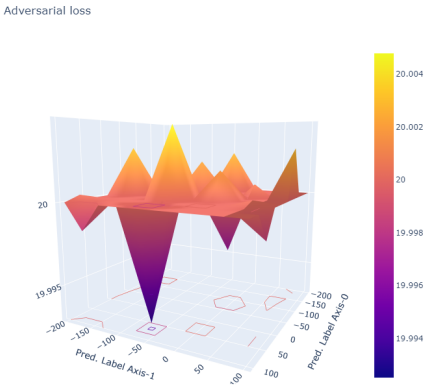
[16] Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein gan. arXiv preprint arXiv:1701.07875.

[17] Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of wasserstein gans. In Advances in neural information processing systems (pp. 5767-5777).

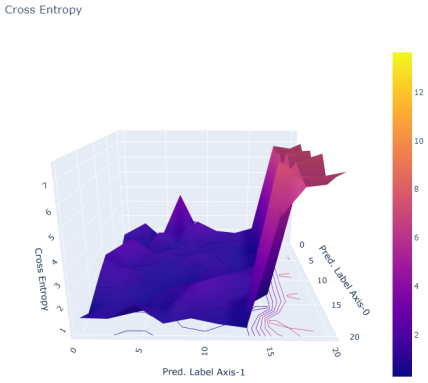
[18] Goodfellow, I. (2016). NIPS 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv:1701.00160.



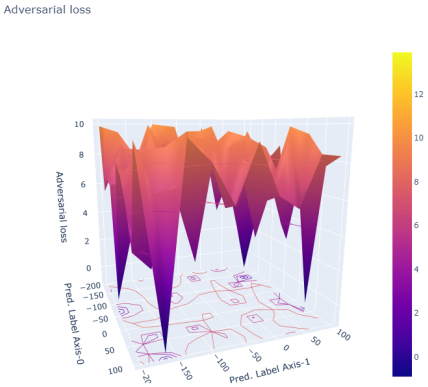

(a) CE Loss landscape at Epoch 0




(b) Adversarial Loss landscape at Epoch 0

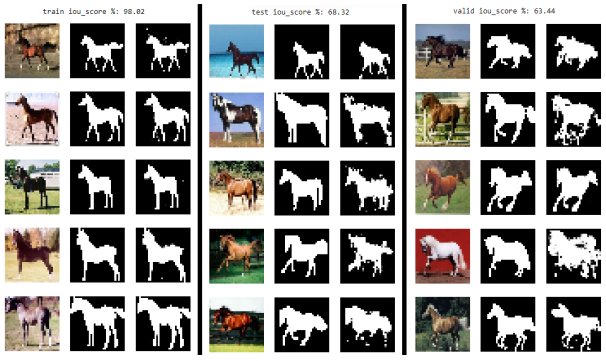



(c) CE Loss landscape at Epoch 100

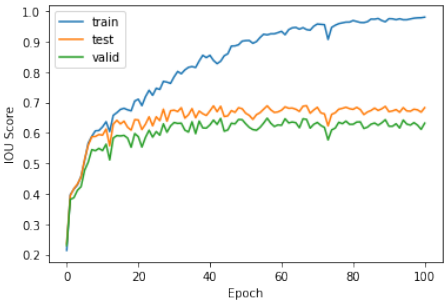



(d) Adversarial Loss landscape at Epoch 100

Figure 10: Loss landscape training progression for Log-loss GAN

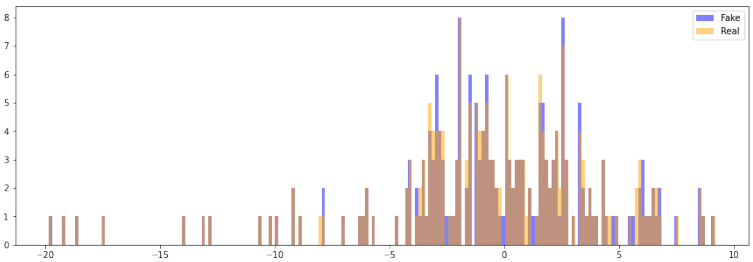



(a) Samples at Epoch 0

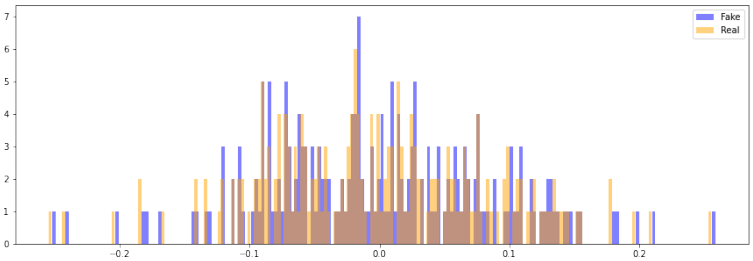



(b) IOU Score progression through training

Figure 11: EB GAN Sample and IOU scores for Train, Test and Valid dataset (from left to right)

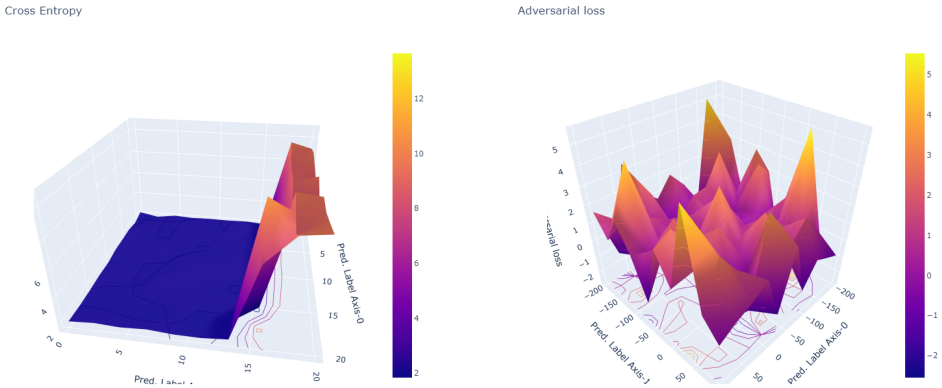



(a) Histogram at Epoch 0



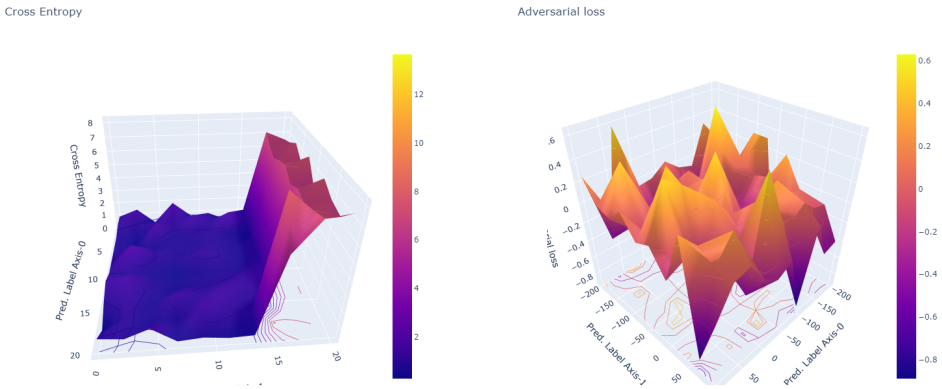

(b) Histogram at Epoch 300

Figure 12: Histograms of Fake vs Real Discriminator Scores For MSE GAN




(a) CE Loss landscape at Epoch 0

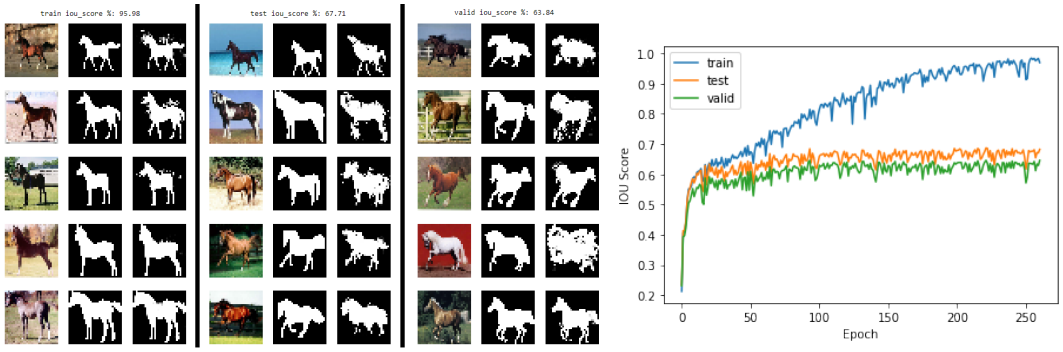
(b) Adversarial Loss landscape at Epoch 0




(c) CE Loss landscape at Epoch 100

(d) Adversarial Loss landscape at Epoch 300

Figure 13: Loss landscape training progression for MSE GAN




(a) Samples at Epoch 0

(b) IOU Score progression through training

Figure 14: MSE GAN Sample and IOU scores for Train, Test and Valid dataset (from left to right)