Kevin Ma - CS287H Final Report

Integrating Human Evaluations into Generative Design Process

1. Introduction

Generative design can be an effective means of enhancing early stage conceptual engineering design processes. For example, engineers can use generated designs to support novel ideation, prevent idea fixation, and expand their realm of design space. In the past, research on generative designs have included using machine learning models to generate 3D designs based on various computational physical and geometrical analysis [1][2]. While these analyses can inform models to generate optimal designs, a design that is fully optimized does not necessarily mean the design is feasible. Additionally, computationally evaluating whether a design is feasible is difficult since feasibility is largely grounded on engineering intuition that takes into account various factors beyond just the geometrical or physical validity of a design. On the other hand, experts, through the use of their engineering intuition, are capable of evaluating whether a design is feasible or not.

In this paper, we propose that experts are better at evaluating the feasibility of a design than a computer could, and we create a framework to integrate expert preferences into the computational generative design loop. We take inspiration from the artificial intelligence (AI) community where recent studies in the past few years have studied different ways of integrating humans into the training process of various AI models [3-5]. We integrate this framework into a generative adversarial network (GAN) model, a commonly used model to generate 3D engineering designs in the mechanical engineering domain [6][7]. During the training process of the GAN, the expert will guide the model's training process by comparing and choosing between a pair of generated outputs. In this report, we opted to use 2D images of celebrity faces instead of 3D engineering designs due to computational limitations and time constraints. Thus, this final report can be viewed as a study on the efficacy of integrating human preferences into generative models.

2. Methods

We divide this section into two parts. In the first two sections, we provide an overview of GANs and Preference Based Reinforcement Learning. In the third section, we introduce the framework that was used in this report.

2.1. GANs

GAN models [8] are a series of generative models via an adversarial network, where a generator *G* generates data based on some noise input and a discriminator *D* takes in real data and, adversarially, discriminates between the generated data by *G* and the real data. The objective function of a GAN is given by equation 1 below:

$$L_{GAN} = E_x[log(D(x))] + E_z[log(1 - D(G(z)))] \quad (1)$$

where $D(x)$ is the discriminator's estimate of the probability that the real data instance, x, is real, E_x is the expected value over all real data, $G(z)$ is the generated output drawn from the random latent variable z, E_z is the expected value over all the random inputs to the generator, and $D(G(z))$ is the discriminator's estimate of the probability that the generated output drawn from the random latent variable z is fake. The discriminator D and generator G plays a minimax game whereby the discriminator D tries to maximize the probability that it classifies reals, $log(D(x))$, and minimizes the probability that D will predict outputs that are fakes, $log(1-D(G(z)))$.

The traditional GAN models have, however, been known to be unstable during training, often resulting in ridiculous outputs. To stabilize the training of the GAN, we use a variation of the traditional GAN, Deep Convolutional Generative Adversarial Networks (DCGAN) [8]. DCGAN [9] introduces a discriminator that contains strided convolutional layers, batch norm layers, and LeakyReLU activations and a generator that contains convolutional-transpose layers, batch norm layers, and ReLU activation layers. Additionally, we opt to use DCGAN as it has been proven to be successful in generating 3D designs of ships [5].

2.2. Preference Based Reinforcement Learning

Integrating human preferences into a deep reinforcement learning model was proposed in [4]. *Christiano et. al.* [4] claims that the reward model in a reinforcement learning algorithm is hard to clearly define, and previous methods such as providing human demonstrations to train an undefined reward model through techniques, such as Inverse Reinforcement Learning [9], are extremely expensive and would require hundreds or thousands of hours of experience. Therefore, they propose to fit a reward function via human preferences while simultaneously training a policy to optimize the current predicted reward function.

Per the framework in [4], their method maintains a policy and a reward function estimate, which will be parameterized by deep neural networks. The policy will produce a set of trajectories based on a set of observations, and the parameters of the policy will be updated by a classical reinforcement learning algorithm to maximize the sum of the predicted reward. A pair of segments from the trajectory will be sent to the human for comparison. Based on the human preferences for the trajectories, a reward function estimate will be optimized via supervised learning to fit the comparisons collected from the humans.

2.3. Proposed Framework

In this paper, we utilize a similar framework to [4] but apply it to the generative design process. In our framework, instead of generating trajectories for humans to evaluate, we generate images for humans to evaluate. In a similar manner to [4], a reward function estimate will be optimized via supervised learning to fix the comparisons collected from the humans. The parameters of the generator will be fine-tuned based on the reward function estimate. Both the generator and the reward function will be parameterized by deep neural networks.

The networks are updated by the following steps:

1. The generator is first pre-trained on existing datasets through the DCGAN model as described in Section 2.1.
2. The pre-trained generator, *G(z)*, takes in two random inputs, z_1 and z_2, which generates 2 designs.
3. Every thousand iterations, the pairs of generated outputs are sent to the humans for comparison.
4. The parameters of the reward function estimate are optimized via supervised learning.

Similar to [4], the reward function is fitted according to a Bradley-Terry model and the Luce-Shepard choice rule [11][12]. We interpret a reward function estimate *r* as a preference-predictor if we view *r* as a latent factor explaining the human's judgment and assume that the human probability of preferring one generated output $G(z_i)$ depends exponentially on the latent reward summed over the length of the training period.

$$P[G(z_1) > G(z_2)] = \frac{exp\Sigma r(z_1)}{exp\Sigma r(z_1) + exp\Sigma r(z_2)} \quad (2)$$

We choose $r(z_i)$ to minimize the cross-entropy loss between the predictions and the actual human labels.

$$loss(r) = -\Sigma l_1 log(P[G(z_1) > G(z_2)]) + l_2 log(P[G(z_2) > G(z_1)]) \quad (3)$$

The labels l_1 and l_2 are determined by whether the human preferred the output generated from the random latent variable z_1 or the random latent variable z_2. Thus, if the human preferred $G(z_1)$ over $G(z_2)$, l_1 would equal 1 and l_2 would equal 0 and vice versa. If the human preferred neither l_1 nor l_2, both values are equal to 0.5.

After the reward function estimate is trained, the pre-trained generator's parameters are updated by running the DCGAN algorithm again; however, this time there is the additional goal of finding the random variable *z* that maximizes the estimated reward from the trained reward function estimator *R(z)*.

$$L_{total} = E_x[log(D(x))] + E_z[log(1 - D(G(z)))] - \Sigma R(z) \quad (4)$$

whereby *R(z)* is the sum of all outputs as a function of the random latent variable z.

3. Results & Discussion

The goal of this report is to investigate the impact of the human preference modeling on the generative process in a GAN model. For the purpose of this report, due to computational and time constraints, we opted to use 2D images of celebrity faces, also known as CelebA [13], for our human preference study instead of using 3D engineering models. As a result, this report could be seen as a proof-of-concept study for this framework, and we hope to use the results of this study to expand the framework to include 3D engineering models.

We train two DCGAN models, one with and without a human preference model. Both DCGANs were run using an Adam optimizer [14] with an epoch of 10, learning rate of 0.0002, mini batch-size of 72, and beta1 of 0.5. Both models are trained on the same dataset, containing 30,047 images. For one experiment, we trained the DCGAN model with a human preference model, as described in Section 2.3, where the expert spends about 30 seconds evaluating which set of images they prefer more. We also have one control experiment where the DCGAN model is trained without any human preferences. More details on the implementation of the DCGAN and human preference model can be seen in the GitHub repository[1].

Figure 1. On the left are the generated images using DCGAN + human preferences and on the right are the generated images using only the DCGAN model.

As seen in Figure 1, qualitatively, there are marginal differences between the images generated using only the DCGAN model and the human preference model. This is due to the fact that the evaluation criteria for which set of celebrity images is preferable for a human is unclear and entirely subjective. In addition, it is also unclear as to whether the difference in the generated image is primarily due to the randomness from the generator, or there is a substantial impact from the trained human preference estimator. In future works, the plan is to utilize 3D engineering designs as training datasets instead of 2D images where the impact of the human preference model will be clearer because the generated output can be judged with more clear criterias of feasibility and usefulness. We now examine the generator and discriminator losses for both experiments, which is shown in Figure 2 below.

[1] https://github.com/kevinma1515/CS287H-project/tree/main

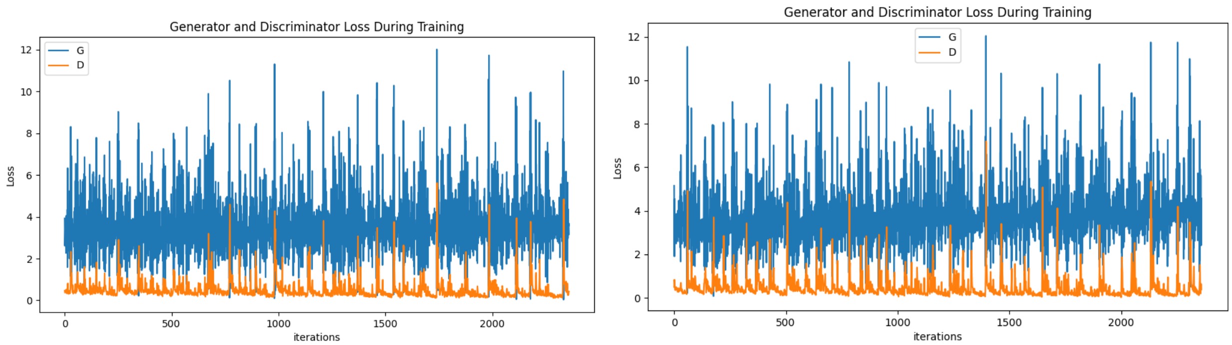

Figure 2. On the left are the generator and discriminator losses using DCGAN + human preference and on the right are the generator and discriminator losses using only DCGAN

Figure 2 above shows similar losses in both generator and discriminator during training for both experimental settings. These results reveal that both models contain similar losses during training, which implies that the human preference model has an insignificant impact on the generator and discriminator's loss values.

4. Conclusion

While this report demonstrates a framework in the use of human preference modeling to guide the generative design process, it contains major limitations in its scope and impact. First, because of time constraint and computational limitations, this report does not utilize 3D engineering design models as its training dataset and opted to use 2D images instead. To address the original research question presented in this paper, future work would need to construct a pipeline that can smoothly visualize generated 3D design outputs for an expert to evaluate during training.

Another major limitation to our framework is the amount of data the expert needs to provide the reward network function estimator during training. The expert has to spend on average 45 minutes to complete the training of the reward network estimator. Noted as a limitation in [4], the efficiency of learning from human preferences could be drastically improved. Finally, another limitation is the method in which the pre-trained generator is being fine tuned using the reward function estimator. The loss function, as described in Section 2.3, is a naive approach to modifying the GAN loss function whereby the reward function estimator acts analogously as a regularizer. The impact of this method is unclear, so other approaches, such as rejection sampling, should be explored.

Overall, the report is an investigation on whether or not human preference modeling could be integrated into the generative design process. In particular, we focus on the integration of human preference modeling into GAN models as these models are widely used in the engineering design community. Successful future improvements on this research project could result in paper publications to design journals, such as Journal of Mechanical Design and Computer-Aided Design, or human computer interaction conferences, such as ACM CHI.

References

[1] Oh, S., Jung, Y., Kim, S., Lee, I., & Kang, N. (2019). Deep generative design: Integration of topology optimization and generative models. Journal of Mechanical Design, 141(11).

[2] Wang, Y., Joseph, J., Aniruddhan Unni, T. P., Yamakawa, S., Barati Farimani, A., & Shimada, K. (2022). Three-dimensional ship hull encoding and optimization via deep neural networks. Journal of Mechanical Design, 144(10), 101701.

[3] Christiano, P. F., Leike, J., Brown, T., Martic, M., Legg, S., & Amodei, D. (2017). Deep reinforcement learning from human preferences. Advances in neural information processing systems, 30.

[4] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems, 35, 27730-27744.

[5] Kazemi, H., Taherkhani, F., & Nasrabadi, N. (2020). Preference-based image generation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 3404-3413).

[6] Khan, S., Goucher-Lambert, K., Kostas, K., & Kaklis, P. (2023). ShipHullGAN: A generic parametric modeller for ship hull design using deep convolutional generative model. Computer Methods in Applied Mechanics and Engineering, 411, 116051.

[7] Chen, W., Chiu, K., & Fuge, M. D. (2020). Airfoil design parameterization and optimization using bézier generative adversarial networks. AIAA journal, 58(11), 4723-4735.

[8] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2020). Generative adversarial networks. Communications of the ACM, 63(11), 139-144.

[9] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint arXiv:1511.06434.

[10] Ng, A. Y., & Russell, S. (2000, June). Algorithms for inverse reinforcement learning. In Icml (Vol. 1, p. 2).

[11] Ralph Allan Bradley and Milton E Terry. Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika, 39(3/4):324–345, 1952. Duncan Luce. Individual choice behavior: A theoretical analysis. Courier Corporation, 2005.

[12] Roger N Shepard. Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. Psychometrika, 22(4):325–345, 1957.

[13] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision (pp. 3730-3738).

[14] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.