How Does Visualization Help People Learn Deep Learning? Evaluation of GAN Lab

Minsuk Kahng*
Georgia Institute of Technology

Duen Horng (Polo) Chau†
Georgia Institute of Technology

Abstract

While a rapidly growing number of people want to learn artificial intelligence (AI) and deep learning, the increasing complexity of such models poses significant learning barriers. Recently, interactive visualizations, such as TensorFlow Playground and GAN Lab, have demonstrated success in lowering these barriers. However, there has been little work in formally evaluating these tools. This paper presents our ongoing efforts in evaluating GAN Lab, an interactive tool designed to help people learn how Generated Adversarial Networks (GANs) works. Through an observational study, we investigate how the tool was used and what users had learned from their usage. Based on the study and our experience in developing and successfully deploying the tool, we discuss future research challenges in the evaluation of interactive educational tools for AI.

Index Terms: Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

With the recent advances in artificial intelligence (AI) and deep learning, a rapidly growing number of people want to learn a variety of new deep learning models. However, the increasing complexity of such models poses significant learning barriers. Recently, interactive visualizations have demonstrated success in tackling this challenge [4, 6, 11, 13, 14]. For instance, TensorFlow Playground [13] allows users to directly manipulate a visualization of neural networks, which has been used to educate employees at Google about deep learning. Furthermore, an increasing number of explorable tools, often called *explorable explanations*, have been developed [4, 14].

While these interactive educational tools have gained popularity and research interest, there has been little work in formally evaluating them. Few works have been published as academic articles [6, 11, 13], some of which include usage scenarios [6]. Evaluation of this new type of tools which focus on educational aspects could be different from that for typical visual analytics tools for interpreting machine learning models [4] or interactive machine learning tools [2].

This paper presents our ongoing efforts in evaluating GAN Lab [6], a recently developed interactive educational tool for Generated Adversarial Networks (GANs), a popular but difficult-to-understand deep learning models. GAN Lab is the first tool designed to help people learn and experiment with complex GAN models in web browsers. We conducted a small observational study to investigate how features in GAN Lab were used and what users had learned from their usage. Based on the study and our development experience, we discuss future challenges in evaluating interactive educational tools, such as how to measure users' level of understanding of machine learning models.

*e-mail: kahng@gatech.edu
†e-mail: polo@gatech.edu

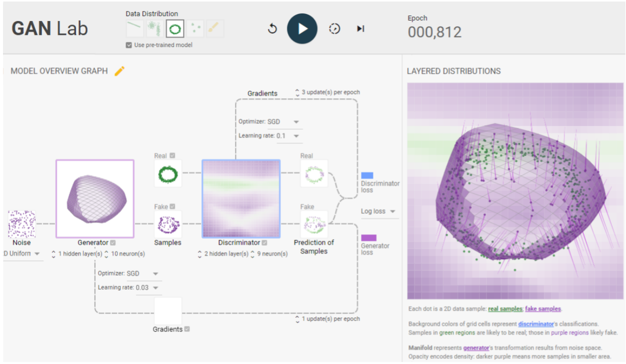

Figure 1: GAN Lab visualizes the structure of GAN models and allows users to interactively train and experiment with the models, helping them actively and playfully learn about GANs.

2 GAN Lab: Interactive Experimentation of GANs

GAN Lab [6], an interactive visualization tool for learning GANs, was designed and developed through a collaboration between Georgia Tech and Google's People+AI Research (PAIR) group; the authors of this paper were part of the team. GAN Lab supports a growing population of people who want to learn deep learning, but had a hard time doing so because of the complexity of modern deep learning models. GANs [3] is a great example of such models. To lower the learning barriers, we built GAN Lab [6].

GAN Lab (Fig. 1) enables users to interactively train a GAN, tweak its hyperparameters, and study how the model responds to generate data distributions. GAN Lab's visualization techniques work in tandem to help people understand complex GAN concepts. The interface tightly integrates a *model overview graph* that summarizes GAN's structure as a graph (Fig. 1 left), selectively visualizing components crucial to the training process; and a *layered distributions view* (Fig. 1 right) that helps users interpret the interplay between the *generator* and *discriminator*, the two key components of GANs.

Deployment. GAN Lab was open-sourced and launched in September 2018 at `https://poloclub.github.io/ganlab/`. It has received significant attention. Within the first year, more than 70,000 people from over 160 countries tried it out.

3 Observational Study

To investigate how GAN Lab's target users (e.g., students aspired to learn about GANs) would use the tool and learn about the models, we conducted a small observational study. This section describes our study design and findings.

3.1 Experiment Design

Participants. Six participants were recruited through our institution's mailing list for those who are interested in machine learning. We pre-screened participants to ensure that they have at least basic knowledge of deep learning and GANs (e.g., taken a deep learning course or at least heard of GANs). Five participants were Ph.D.

students who had taken a deep learning course, and one was an undergraduate student who had research experience. They self-reported their level of knowledge on deep learning, with an average score of 3.3 on a scale of 0 to 5 (0 being "no knowledge" and 5 being "expert"); and that on GANs with an average score of 2.5 (on the same scale). No participant has used or heard about GAN Lab before.

Procedure. The study was conducted through BlueJeans video conferencing. After the participants signed their consent forms electronically, they were provided a 5-minute overview of GANs, followed by a 5-minute tutorial of GAN Lab, which described its visualizations and features. After that, the participants freely explored using GAN Lab on their computer's web browser. They were asked to think aloud and share their computer screen with us during the study. They could ask for questions when necessary. After they used the tool, the participants were asked to fill out questionnaires. The study took about 50 minutes, and each participant was compensated with an Amazon $15 gift card for their time.

3.2 Key Findings

Rapid hypothesis testing. Among the features of GAN Lab, many participants particularly liked the one for dynamically adjusting hyperparameters while a model was being trained. This feature enabled them to form hypotheses based on prior experience in machine learning and rapidly test them using GAN Lab. For example, one participant increased the learning rate (using its drop-down menu) to test if it helps speed up the training. Another participant said "*I really liked the features of the hyperparameter tuning [...], and learning all the different hyperparameters that can affect them are making me think of different ways to optimize GANs.*" This capability for rapid hypothesis testing in GAN Lab is not possible in conventional deep learning workflows because they often require retraining the model each time a user adjusts a hyperparameter.

Building intuition through dynamic experiments. The ability to adjust hyperparameters in GAN Lab also helps users build intuition about the behaviors induced by the model's training process. One important characteristic of GANs is the dynamic interplay between the two components: generators and discriminators. A participant said "*[the] ability to change training parameters such as number of updates on the fly was nice. It really helps you build intuition to see how the discriminator and generator interact.*" One usage pattern participants particularly liked was updating either the generator or discriminator while disabling the update of the other. By default, the training process alternates between the generator and discriminator (in each iteration), so it can be hard for novices to understand their individual contribution to the training progress. By disabling one of them, users can more easily observe how each component works and how the model reaches an equilibrium that balances the two components.

Validating knowledge from literature. Participants who are familiar with the literature of deep learning and GANs found GAN Lab useful for validating knowledge they acquired from research articles. For example, one participant remembered that GANs would often encounter the problem called *mode collapse*, especially when a distribution contained disjoint modes [7]. This participant was interested in reproducing this phenomenon by training a model with such a distribution. He also wanted to use a different loss function that might mitigate this issue, as suggested in the literature. This observation suggests that interactive tools like GAN Lab may help not only novices learn the basic concepts of models, but also researchers and practitioners validate knowledge they learned from the literature, which could help them build trust in the model's training process.

Beginners need further guidance. We observed that participants less familiar with GANs needed more guidance to help them fully enjoy the tool. Some were not sure about what to try. One said "*helpful to [provide descriptions] of what GANs training scheme "works" and what "doesn't work."*" Although we wanted users to self-discover relationships between hyperparameters and results by actively playing with the tool, it might be beneficial for us to also provide step-by-step exercises that would guide users' experimentation, similar to how TensorFlow Playground has been integrated into Google's machine learning course material on the web [1]. The course includes a series of exercises which learners can follow. For example, in the chapter on learning rates, learners are asked to try different learning rates and compare the results.

4 Discussion: Measuring Understanding

Our observational study is an early step in understanding how people may learn deep learning through interactive education tools. There remain many challenges in designing controlled experiments to further such evaluation efforts. One important challenge is the choice of dependent variables that measure a user's level of the understanding in machine learning (ML) models, similar to the use of task completion time for evaluating information exploration tools. We briefly discuss this challenge here.

Studies conducted in computer science education research and those for evaluating algorithm visualizations (in early 2000s) typically included pre- and post-study tests that sought to measure participants' conceptual or procedural knowledge (e.g., what is the algorithm's time complexity, what would be the next state after '17' is inserted) [5]. However, test questions suitable for simpler, deterministic algorithms may not generalize to modern ML models that are often complex and probabilistic.

Thus, it would be a valuable effort to develop new ways to evaluate the educational effectiveness of interactive tools for ML. Below we present a few ideas. First, the computer science education literature has developed several methods, such as analyzing mental models or measuring self-efficacy [8, 12], and we can draw inspirations from them. Next, inspired by how visual analytics tools are evaluated [10], studies may be designed to analyze if participants discovered new insights on ML models. In addition, since the primary goal of ML learners is often in developing models for real data, it could be helpful to design studies that assess if users are able to implement models with high accuracy.

Lastly, we wanted to note that the level of understanding is not the only dependent variable in evaluating educational tools. Another important factor to measure is the learners' engagement level [9]. A high level of engagement (e.g., spending more time and efforts) often indicates that users enjoy the tool and may likely learn more through the usage. To investigate if GAN Lab users are actively engaged, we have been collecting anonymous usage log (e.g., buttons users clicked) from our deployed website and plan to analyze them.

5 Conclusions

This paper presents our ongoing efforts in evaluating GAN Lab, an interactive tool for learning a popular, complex deep learning model. Through an observational study, we found that GAN Lab helps users learn about GANs by interactively training models using multiple features, such as dynamic adjustment of hyperparameters. Based on the study, we discussed further evaluation challenges in designing controlled experiments and our plan to perform an analysis of usage log to measure the users' engagements. We believe tools like GAN Lab have a huge potential for promoting people's understanding of machine learning models, and hope our work will inspire more research, development, and evaluation of such tools.

Acknowledgments

We thank Martin Wattenberg and Fernanda Viégas for their advice. This work is supported by a Google PhD Fellowship.

REFERENCES

[1] Neural networks: Playground exercises, machine learning crash course. `https://developers.google.com/machine-learning/crash-course/introduction-to-neural-networks/playground-exercises`, 2019. Accessed: 2019-09-07.

[2] N. Boukhelifa, A. Bezerianos, and E. Lutton. Evaluation of interactive machine learning systems. In *Human and Machine Learning*. Springer, 2018.

[3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, 2014.

[4] F. Hohman, M. Kahng, R. Pienta, and D. H. Chau. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE Transactions on Visualization and Computer Graphics*, 25(8), 2019.

[5] C. D. Hundhausen, S. A. Douglas, and J. T. Stasko. A meta-study of algorithm visualization effectiveness. *Journal of Visual Languages & Computing*, 13(3), 2002.

[6] M. Kahng, N. Thorat, D. H. Chau, F. B. Viégas, and M. Wattenberg. GAN Lab: Understanding complex deep generative models using interactive visual experimentation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 2019.

[7] L. Metz, B. Poole, D. Pfau, and J. Sohl-Dickstein. Unrolled generative adversarial networks. In *5th International Conference on Learning Representations (ICLR)*, 2017.

[8] T. Naps, S. Cooper, B. Koldehofe, C. Leska, G. Rößling, W. Dann, A. Korhonen, L. Malmi, J. Rantakokko, R. J. Ross, J. Anderson, R. Fleischer, M. Kuittinen, and M. McNally. Evaluating the educational impact of visualization. *ACM SIGCSE Bulletin*, 35(4), 2003.

[9] T. L. Naps, G. Rößling, V. Almstrum, W. Dann, R. Fleischer, C. Hundhausen, A. Korhonen, L. Malmi, M. McNally, S. Rodger, and J. A. Velázquez-Iturbide. Exploring the role of visualization and engagement in computer science education. *ACM SIGCSE Bulletin*, 35(2), 2003.

[10] C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3), 2006.

[11] A. P. Norton and Y. Qi. Adversarial-Playground: A visualization suite showing how adversarial examples fool deep learning. In *IEEE Symposium on Visualization for Cyber Security (VizSec)*, 2017.

[12] V. Ramalingam, D. LaBelle, and S. Wiedenbeck. Self-efficacy and mental models in learning to program. *ACM SIGCSE Bulletin*, 36(3), 2004.

[13] D. Smilkov, S. Carter, D. Sculley, F. B. Viégas, and M. Wattenberg. Direct-manipulation visualization of deep networks. In *ICML Workshop on Visualization for Deep Learning*, 2016.

[14] M. Wattenberg, F. Viégas, and I. Johnson. How to use t-SNE effectively. *Distill*, 2016.