GANs Unplugged

Patrick Virtue
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, Pennsylvania 15206
pvirtue@cs.cmu.edu

Abstract

With the influx of deepfake and style transfer technology in today's news and social media, everyone is told that these applications are powered by artificial intelligence and deep learning, but too often the explanation of how it works goes no further. Rather than waiting until second semester of computer science grad school to learn about generative adversarial networks (GANs), we propose a classroom activity to introduce GANs to secondary school students.

Our GANs Unplugged activity steps outside a traditional classroom environment and forms groups of students that physically act as the various components of a GAN. Students in the Real and Generator (i.e. Fake) groups both draw sketches of animals on index cards, but only the Real group has the secret, specific instructions on how to draw their animals. A third Discriminator group tries to determine if each card it receives came from the Real group or the Fake group. The Discriminator group also makes a single feedback mark on the drawing to suggest how that drawing could be more "real" before passing the card back to the group that drew it. As this game progresses, the Discriminators get better at determining real and fake, that is until the Generators adapt and learn to draw more and more realistic "fake" images.

We describe the details of GANs Unplugged and present our experience running this activity at three different artificial intelligence summer camps for high school students.

Introduction

In 2014, Goodfellow et al. introduced Generative adversarial networks (GANs) (Goodfellow et al. 2014) as an effective means to generate unique data that can be indistinguishable from those in the training dataset. This has lead to a surge in breakthrough technology in tasks such as image reconstruction, style transfer, and both benign and malignant fake data generation. While it takes quite a bit of training in machine learning to understand the details in the original GAN paper, we argue that the fundamental concepts behind GAN technology may be presented to secondary school students with little or no artificial intelligence training.

We present our GANs Unplugged activity where students physically act out the different components of GANs while playing an engaging game to ultimately sketch realistic-looking fake images, Figure 1.

We aim to achieve the following student learning outcomes:

- Explain how the different components of GANs work together to generate realistic fake data.
- Describe how learning to detect real versus fake is an integral part of GANs ability to create fake data.
- Discuss societal impacts of successful fake data generators.

These learning outcomes are achieved through three stages of the GANs Unplugged activity: preparing the students, running the game itself, and debriefing the students.

Related Work

Our GANs Unplugged activity is part of the larger Computer Science Unplugged initiative that allows students to step away from the computer, and potentially out of the classroom, to focus on computational thinking without being distracted by the mechanics of coding or mathematics (Bell, Witten, and Fellows 1998; Bell 2020; Bell and Vahrenhold 2018). Unplugged activities also have the added benefit of broadening participation in computing, as they may be used in schools with limited resources for educational technology (Cortina 2015). Unplugged activities have been developed for artificial intelligence topics. For example, the "Intelligent Piece of Paper" exercise engages students in an activity and subsequent discussion about the intelligence of a piece of paper containing logic to play a perfect game of tic-tac-toe (Curzon 2016).

The unplugged activities that are most closely related to GANs Unplugged are the neural network unplugged exercises within the Artificial Intelligence unit of the Exploring Computer Science high school curriculum (Clarke et al. 2020). In their Forward Propagation activity, students act as different neurons in the network, passing messages to student-neurons in later layers, and ultimately classifying the contents of the input image. Linder, et al. describe a similar forward propagation exercise as part of their series of AI unplugged activities (Lindner, Seegerer, and Romeike 2019).

In addition to being an unplugged activity, our proposed exercise engages students in a discussion about the societal impacts of GAN technology. This complements related AI education efforts to teach primary/ middle school students

about AI ethics and teach high school students about the importance of fairness and explainability of AI systems (Ali et al. 2019; Payne 2020; Alonso 2020).

Placement in Curriculum

We propose that GANs Unplugged will function well as a stand-alone exercise or as part of both middle school and high school AI curricula.

Stand-alone: The GANs exercise is designed to begin with a discussion to prime the students about where this technology appears in everyday life. Because this discussion and the subsequent exercise doesn't require any prerequisite knowledge of AI, programming, or even mathematics, the activity works well as stand-alone exercise for middle school or older students.

Middle school: Sabuncuoglu recently proposed a middle school curriculum that includes high-level components about "How do computers see?" and "How do computers hear?"(Sabuncuoglu 2020). GANs Unplugged would allow instructors to extend this curriculum with an additional component, potentially titled, "How do computers create?". The physically engaging aspects of our activity are well suited for the busy bodies in this age group.

High school: As shown in our results, GANs Unplugged is a natural part of a high school-level AI course or camp, where they have already completed units on machine learning, perhaps regression, classification, and neural networks. During the debrief discussion, the instructor can help the students to make detailed connections between the human GAN components and neural network concepts, such as backpropagation. With more time in the curriculum, instructors could even extend the discussion to talk about how GANs could learn to draw a specific image, e.g. deep fakes and social media image filters, with image-to-image conditional adversarial networks (Isola et al. 2017).

GANs Unplugged Activity

Preparing the Students

GANs Unplugged first begins by introducing students to the real-world applications of GANs, many of which they may have already seen on social media or through apps on their phones. A quick demonstration of deepfake videos and style transfer is a fun way to immediately attract the students' attention. Additional applications, such as transferring autonomous vehicle simulation videos to the real world (Pan et al. 2017) or medical image reconstruction (Yang et al. 2017), can show the students the breadth of possibilities for GANs technology.

This initial presentation should start to build the students' curiosity about how this technology is possible as well as prime them with questions about whether these GANs are beneficial, potentially harmful, or both.

Activity: Student Groups

A total of 20-40 students are first divided into four groups: Real (R), Generator/Fake (G), Discriminator (D), and Wiring (W). Students may be divided evenly among these four groups, though it is often helpful to shift a few students from the Real to Generator group.

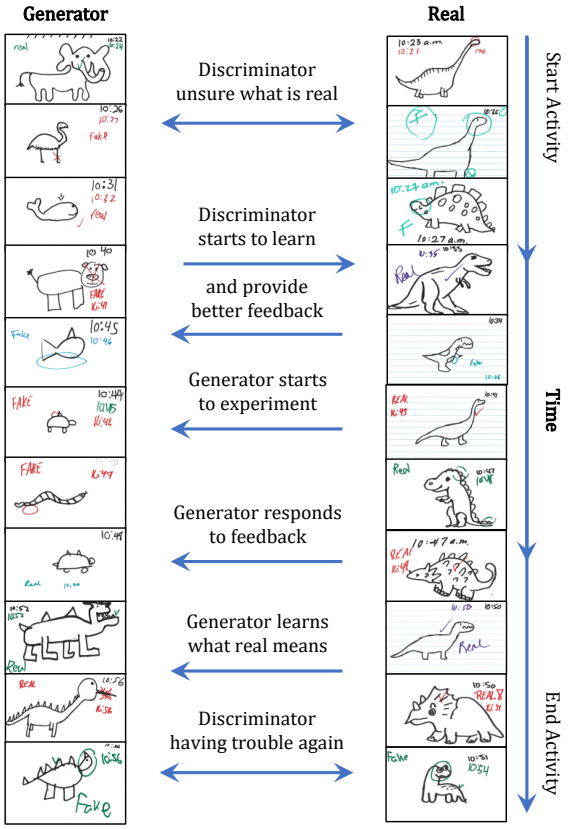

Figure 1: Progression of drawings from Generator and Real student groups, showing how the human GAN gradually learns to generate realistic fake images (dinosaurs, in this case). These are just a few of the hundreds of drawings created during the GANs Unplugged activity.

Activity: Room Setup

Ideally, the activity takes place in two adjacent classrooms as shown in Figure 2. One big room, e.g. a gym, cafeteria, or even outdoors, will work, provided there is some form of partition between groups. Specifically, group D should not be able to see whether cards come from G or R, and groups D and G should not be able to hear each other.

Activity: Materials

Necessary materials include:

- Table/desks for the Real, Generator, and Discriminator groups to write on
- Black markers for the Real and Generator groups, colored markers for Discriminator group
- 200-300 index cards to draw on

Activity: Instructions for Students

The text in this section contains instructions that can be printed and handed to the students in addition to the feed-

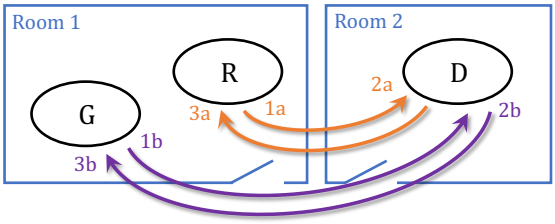

Figure 2: Example layout of student groups in two adjacent rooms. The Wiring group begins by picking up a drawing from either the Real group (R), 1a, or the Generator group (G), 1b. They carry that card to the Discriminator group (D) in the adjacent room, 2a/2b. After waiting for the Discriminator group to add their label and feedback markings, the Wiring group carries the annotated card back to the original group, 3a/3b.

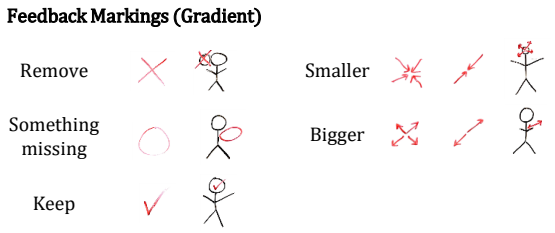

Figure 3: Legend of feedback marks given to all groups. The Discriminator group will draw these marks on all images, and the Generator group will use this feedback signal to make their drawings more realistic.

back marking guide, Figure 3.

"Today, we are going to learn about how Generative Adversarial Networks (GANs) work by playing a game! You will be divided into groups (Real, Discriminator, Generator, and Wiring), and each group must follow different instructions as written below. The Real group will be given a secret instruction to draw a particular animal, and our goal today is to guess what this secret animal is. Let's play!"

For the Real group: "The instructors will tell you to draw a specific animal. As the Real group, you will be the only group who knows what this real instruction is. Keep this information a secret! It's very important that the other groups (except the Wiring) do not know what you are drawing. You will draw your animals on index cards, and pass them to the Wiring who will then pass them to the Discriminator group. Write a time stamp on each drawing that you finish."

For the Generator group: "Your group will not receive the real information on what animal the Real group is instructed to draw, but your role is to generate guesses. You will be given index cards to draw on. Start by drawing any animal that you like, write a time stamp on it, and pass it to the Wiring who will pass it to the Discriminator group. Afterward, the Wiring will give your picture back to you, with feedback on which part of the picture should be changed. Use this information to guess what the real animal should be! Talk to each other! Don't forget to write a time stamp on each drawing that you finish."

For the Discriminator group: "As the Discriminator, your role is to try to guess which drawings are real or fake. You will receive index cards from the Wiring, and you will have to guess which drawings are from the Real group and which are from the Generator group. If you think a picture is real, write 'Real' on it and put a tick mark on which part of the picture you want to keep. If you think a picture is fake, write 'Fake' on it, and leave just one feedback mark on that card according to the feedback marking legend given to you. Talk to each other to guess what the real animal should be! When you're done, pass the card back to the Wiring, who will tell you if you get your guess right. Don't forget to write a time stamp on each card that you give back to the Wiring."

For the Wiring: "As the Wiring, your role is to pass the information between the data generator groups (Real and Generator) and the Discriminator group. Your role is very important! Be careful not to give the cards to the wrong group, so that we can keep guessing what the real picture should be. You should also be careful not to reveal what the real animal is to the Discriminator group! The Discriminators will tell you their guess of whether a drawing is real or fake, and you can only tell them if they were right or wrong, without telling them what animal is right."

Activity: Execution

The activity begins with the instructor whispering to the Real group the secret attributes that define the real images, e.g. "the real animals should all be dinosaurs". As images are drawn, passed between groups, and marked up, instructors can monitor the progress. Initially, instructors will be focused on ensuring the students understand and are following the instructions of the game. As the game progresses, the instructors should encourage discussion within the Generator and Discriminator groups to share what they've seen and what they've learned so far. Instructors may need to encourage the Generator group, in particular, to think critically about what images they might want to draw next to better understand the meaning of "real". The game continues for about 30 minutes and may be stopped once the Generator group has been consistently drawing realistic fake images.

Instructors should encourage the Wiring group to quietly observe the progression of the game as the Generator and Discriminator groups begin to learn. The Real group, while an important piece in the GAN, unfortunately, doesn't have to make any critical decisions. To keep this group engaged and to allow them to learn from the whole process, instructors should encourage one member at a time from the Real group to quietly walk around the two rooms and observe the entire process in the works.

Ideally, there would be time for a second round of the game where the groups are rotated and the secret real attributes are changed for the new Real group.

Debrief Discussion

Viewing/analyzing the progression of images: After the game has finished, ask the Real and Generator groups to each roughly sort the cards returned to their group by the

timestamps written on the card. Give all students a few minutes to walk past these drawings, noting the changes in the two groups as time progressed. Begin the initial debrief discussion by asking the students what they observed, both in these images and in the group discussions. Students and instructors can engage in a discussion about the stages of the GAN learning as highlighted in Figure, 1.

If time permits, the discussion can be extended by posing the question, "what might have happened if we let this game continue for another hour or longer?" Potentially, the Discriminator could start to recognize the particular style of dinosaurs that the artists in the Real group tend to draw. The Discriminator group will then start to successfully determine real from fake again. Would the Generator group then eventually to learn to draw that particular style?

Making connections to artificial GANs: Now that the human GAN has finished learning, describe the components of an artificial GAN, and help students draw connections from the various GAN components to the four human groups. Instructors might consider launching this discussion by asking students, "which of the four human groups were learning?" For more advanced groups, e.g. more senior high school students towards the end of an introductory AI curriculum, you can draw specific connections to what they have learned about neural networks; for example how the Wiring group relates to backpropagation.

Discussion about GANs and deepfakes in the real world: It is important to save time at the end to hold a conversation about applications in the real world. Where can GANs be beneficial? What do we have to be cautious about? This discussion is usually not difficult to start as students are typically primed to ask questions like, "what happens if we have fake videos that even a computer can't detect as fake?"

Results

We ran the GANs Unplugged activity as part of three different high school AI summer camps: one at our university and two at local Boys & Girls Clubs of America. We observed that all camps were able to successfully complete the activity, generating fake images within 20 to 40 minutes, all following the general progression shown in Figure 1. For all three summer camps, the GANs exercise was part of an introductory AI curriculum. Prior to the GANs exercise in the curriculum, students had learned about various AI tasks and techniques including regression, classification, and a brief introduction to neural networks.

To test the GANs Unplugged as a stand-alone activity outside the context of an intro AI curriculum, we ran this exercise as one of a series of evening STEM sessions for local middle school students at our university. Despite having no prior exposure to AI education, this younger age group successfully completed the GAN simulation, generating fake images within 30 min.

Both the high school groups and the middle school group were extremely engaged in the debrief discussion, asking and answering questions until our session time frame expired.

We look forward to sharing GANs Unplugged with a larger audience, helping students around the world gain insight into the inner workings of this impressive AI technology.

References

Ali, S.; Payne, B. H.; Williams, R.; Park, H. W.; and Breazeal, C. 2019. Constructionism, Ethics, and Creativity: Developing Primary and Middle School Artificial Intelligence Education. In *International Workshop on Education in Artificial Intelligence K-12 (EDUAI'19)*.

Alonso, J. M. 2020. Teaching Explainable Artificial Intelligence to High School Students. *International Journal of Computational Intelligence Systems* 13(1): 974–987.

Bell, T. 2020. CS Unplugged. https://csunplugged.org/. Accessed: 2020-12-18.

Bell, T.; and Vahrenhold, J. 2018. CS Unplugged—How Is It Used, and Does It Work? In *Adventures between lower bounds and higher altitudes*, 497–521. Springer.

Bell, T. C.; Witten, I. H.; and Fellows, M. 1998. *Computer Science Unplugged: Off-line activities and games for all ages*. Citeseer.

Clarke, B.; Mendelson, M.; McOwan, P.; Curzon, P.; and Lane, R. 2020. AI + Ethics Curriculum for Middle School. https://www.exploringcs.org/for-teachers-districts/artificial-intelligence. Accessed: 2020-12-18.

Cortina, T. J. 2015. Reaching a broader population of students through" unplugged" activities. *Communications of the ACM* 58(3): 25–27.

Curzon, P. 2016. The Intelligent Piece of Paper Activity. https://teachinglondoncomputing.org/resources/inspiring-unplugged-classroom-activities/the-intelligent-piece-of-paper-activity/. Accessed: 2020-12-18.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative AdversarialNets. In *Advances in Neural Information Processing Systems*, 2672–2680.

Isola, P.; Zhu, J.-Y.; Zhou, T.; and Efros, A. A. 2017. Image-to-image Translation with Conditional Adversarial Networks. In *Proceedings of the IEEE Conference on Computer vision and Pattern Recognition*, 1125–1134.

Lindner, A.; Seegerer, S.; and Romeike, R. 2019. Unplugged Activities in the Context of AI. In *International Conference on Informatics in Schools: Situation, Evolution, and Perspectives*, 123–135. Springer.

Pan, X.; You, Y.; Wang, Z.; and Lu, C. 2017. Virtual to Real Reinforcement Learning for Autonomous Driving. *arXiv preprint arXiv:1704.03952* .

Payne, B. H. 2020. AI + Ethics Curriculum for Middle School. https://www.media.mit.edu/projects/ai-ethics-for-middle-school/overview/. Accessed: 2020-12-18.

Sabuncuoglu, A. 2020. Designing One Year Curriculum to Teach Artificial Intelligence for Middle School. In *Proceedings of the 2020 ACM Conference on Innovation and Technology in Computer Science Education*, 96–102.

Yang, G.; Yu, S.; Dong, H.; Slabaugh, G.; Dragotti, P. L.; Ye, X.; Liu, F.; Arridge, S.; Keegan, J.; Guo, Y.; et al. 2017. DAGAN: Deep De-aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI reconstruction. *IEEE Transactions on Medical Imaging* 37(6): 1310–1321.