Accessibility of Deepfakes

Andrew L Collings

Old Dominion University

Abstract

The danger posed by falsified media, commonly referred to as deepfakes, has been well researched and documented. The software Faceswap to was used to swap the faces of two politician (Joe Biden and Donald Trump). The testing was performed using an affordable consumer GPU (an AMD Radeon RX 570) over 100,000 iterations. The process and results for the two attempts with the best results (and largest differences) were recorded. The result was ultimately unconvincing, while the software was able to recreate the facial structure the lighting and skin tone did not blend at all.

Accessibility of Deepfakes

Consistent with Gordon Moore's prediction, silicon transistor counts have tended to double every 18 to 24 months. This exponential growth has helped to usher in technological advances at an unprecedented rate. As a result, the average consumer has access to an incredible amount of computational power. An entry level graphics card, such as an AMD Radeon RX 570 (AMD, n.d.), can perform over 5 trillion floating point operations (add, subtract, multiply, store, etc.) per second. A mid-range model like and Nvidia RTX 2070 Super (Moass, 2019) has hardware optimized for deep learning that allow it to perform up to 72 trillion operations per second. As such modern graphics cards now deliver a level of performance that would have required a supercomputer 20 years ago.

The methods used to generate 3D graphics on a computer require lots of calculations to be performed at once. Consequently, GPUs (graphics processing units) are engineered to perform a large number of calculations in parallel rather than operate at the highest possible frequency. The algorithms used for machine learning also highly parallelized and as such, tend to perform quite well on a GPU. The fact that many machine learning algorithms are open source and can be used on commodity hardware makes the field accessible and pushes the science forward. Unfortunately, the technology is just as accessible to bad actors.

Deepfake, a portmanteau of deep learning and fake, is an image or video in which a person is replaced with someone else's likeness. The technology has some practical applications, such as in film to make an actor appear as they did when they were younger. It has an equal number of nefarious applications such as revenge pornography, disinformation, etc.

A great deal of research has been done into deepfakes: the threats they pose, the relevant case law, even some of the possible upsides. The current research does not, however, address just how approachable this technology is. Especially given the computational power now available with commodity hardware it makes sense to quantify the time and effort required to create a convincing false video clip.

Method

Using commodity hardware (see Figure 1), leverage Faceswap software to change the subject of a 30-second video clip. Document the process of configuring the software as well as the time required to extract sample data and train the neural network that produces output. As the concept of "convincing" differs from person to person the resulting output will included in addition to subjective findings. The model will continue to be iterated on so long as the result changes substantially or until the research period is exhausted.

Figure 1. *Configuration of Test System*

CPU:	AMD Ryzen 3600
RAM:	16GB DDR4-3200
GPU:	AMD Radeon RX 570 8GB
Operating System:	Windows 10 Professional Build 18363.900
GPU Driver Version:	19.50.29.27-200421a-354308E-RadeonSoftwareAdrenalin2020

Results

The software used to perform the tests is free at Faceswap.dev and is open source (github.com/deepfakes/Faceswap). The website also provides comprehensive documentation on using Faceswap and details on all the configuration options (torzdf, 2019). The installation is completely graphical and requires minimal user intervention beyond selecting graphics card type and destination directory. The installer automatically performs all the tasks typically required for open source artificially intelligence such as installing python, creating an isolated environment, and installing libraries. Launching Faceswap using the desktop shortcut created during setup presents the user with the GUI (Figure 2).

The subjects chosen were Donald Trump (NBC News, 2020) and Joe Biden (ABC News, 2020) as a copious amount of training data exists for both. Images were extracted using the S3Fd detection plugin and the Fan aligner plugin as these were both suggested by the documentation. An output image size of 256x256 pixels was used as this is the maximum supported image size of the recommended training algorithm. Initial attempts were completed without a masker plugin (which maps the face and constrains the calculations to that area) and the final attempt used the Vgg-Obstructed plugin because Donald Trump's face is partially obstructed by a microphone in some scenes.

Face extraction on a 30-second clip took just under six minutes for each subject. The faces without a mask captured part of the background so extraction was performed again with the plugin and completed in 12:06 for Biden and 12:53 for Trump. The 30-second clip produced no false positives (items detected as faces that were not) or misaligned faces. Using the originally planned 2-minute clips (more on this later), Biden had 19 misidentifications and Trump had 26. Each 30-second clip yielded 900 usable images for use as training data.

Training was completed using the Original trainer plugin with a batch size of 64 using the center 68.25% of the image (to focus on the face) over 100,000 iterations. The study was initially planned to use 2-minute clips, but this proved to be impractical as training the model took over a week. The training process for the 30-second clips used completed in 2:04:46:27. The complete result is detailed below (Figure 4).

Figure 2. *Faceswap GUI*

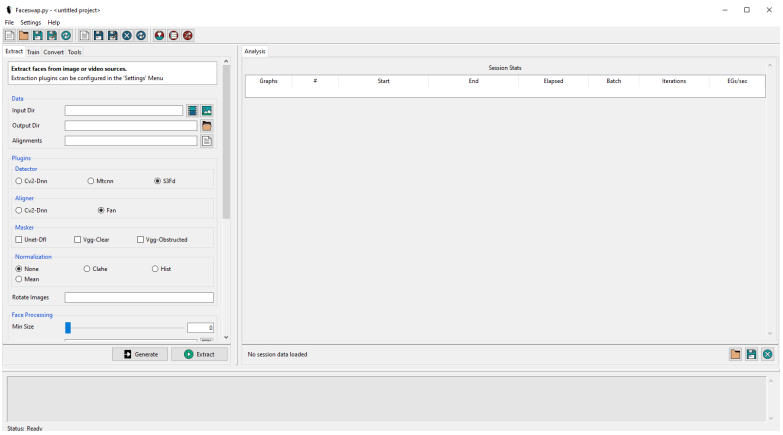
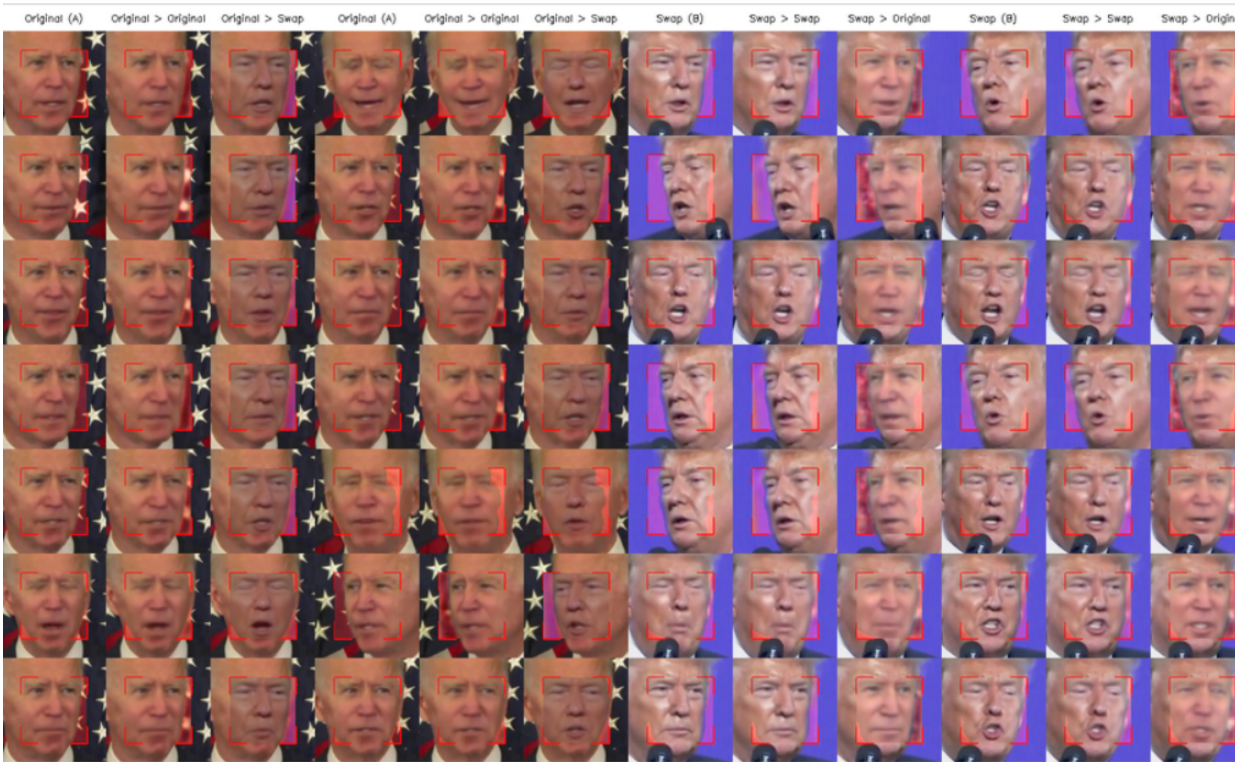


Figure 3. *Background Capture*

Figure 4. *Training Output*



Conclusion

The initial aim of this research was to demonstrate how dangerously accessible image manipulation through artificial intelligence is. It is clearly possible given abundance of examples one can find on the internet. The Faceswap software is also surprisingly intuitive. The plurality of open source artificial intelligence tools are text based, require users to manually create runtime environments, and have very technical documentation. Conversely, Faceswap offers a completely graphical experience (though the command line is still accessible), is largely automated, and has comprehensive documentation (with pictures). Despite this, the results don't seem to support the original hypothesis.

The quality of the output improved substantially from the run without a mask (Figure 3) to the run where a mask was utilized (Figure 4). Especially in the final run, facial features map quite nicely but skin tone and lighting to not match at all. This makes it readily evident that the image has been manipulated. It is possible further improvement could be gained by using more training data from a larger variety of sources. Doing so would, however, increase the amount of time required to train the model (already over 2 days) and require considerably more time and effort to locate and extract said data. Audio also warrants further explorations but was completely outside the scope of this research.

Serendipitously, a popular YouTube channel released a video about deepfakes (Linus Tech Tips, 2020) as this report was being compiled. Their team attempted to create a video with Linus that he wasn't in. Despite having the subject's participation, access to countless hours of high-resolution training data, a professional actor with a similar build, a professional production team, and access to a company specializing in AI audio generation they were still unable to produce a video that most would find convincing. Technology grows at an astonishing pace and further research will be required as the algorithms change and improve but for the time being the applications appear to be more limited than theorized.

References

ABC News. (2020, June 2). Biden addresses nationwide Floyd protests. YouTube. https://www.youtube.com/watch?v=wfiZy4xvhu4

Linus Tech Tips. (2020, July 8). I can safely retire now. YouTube. https://www.youtube.com/watch?v=G0z50Am4Uw4

Moass, Dominic. (2019, July 2). Nvidia RTX 2070 SUPER Founders Edition 8GB Review | KitGuru. KitGuru. https://www.kitguru.net/components/graphic-cards/dominic-moass/nvidia-rtx-2070-super-founders-edition-8gb-review/

NBC News. (2020, June 23). Trump Speaks To Arizona Youth | NBC News. YouTube. https://www.youtube.com/watch?v=fJCIk4OT7EE

Radeon™ RX 570 | Advanced HD Gaming Graphics Card | AMD. (n.d.). AMD. Retrieved 2020, July 15, from https://www.amd.com/en/products/graphics/radeon-rx-570

Torzdf. (2019, September 29). [Guide] Training in Faceswap [Online forum post]. Faceswap. https://forum.Faceswap.dev/viewtopic.php?t=146