NAME: **Ricardo Gonzalez**

PARENTS' NAMES **Ricardo and Maria Gonzalez**

HOMETOWN: **Greenwood, Indiana**

CAREER OBJECTIVE**: I want to work for a consulting firm within the world of Cybersecurity**

BIOGRAPHY: **I am a first-generation university student with a vast interest in computers and computer systems. The security of these systems has been of great interest to me, since I begin my journey at Purdue. I want to assure that all students, friends, and family know the dangers of improper computer security. I hope to one day be able to work as a computer security consultant in order to allow companies to not have to worry about their cybersecurity needs.**

FACULTY MERF SPONSOR: **Dr. Robert M. Stwalley III, Assistant Clinical Professor of Agricultural and Biological Engineering**

GOAL OF THE WORK: **I was trying to discover a way to block unsolicited images that were sent to a person's cellular device using Artificial Intelligence.**

PERSONAL STATEMENT ABOUT THE LESSONS LEARNED FROM THIS EXPERIENCE: **This was the second individual research assignment that I had undertaken. I was able to use the experience and knowledge from my first research experience to make this one far better. I learned how Artificial Intelligence can be used in order to detect images that may be potentially harmful to people. I also learned that when working on solo projects, the project does not have to be done all by one's self. You should, in fact, ask for help when there are areas of a topic that you are unsure about. These lessons are ones that I will take into my next experience in order to make the next one better and even more seamless than the previous one.**

Artificial Intelligence in Cybersecurity

By Ricardo Gonzalez

Abstract

Keeping unsuitable web content from young eyes is a challenge, given the wide-open environment of the internet. Research was conducted into selecting an Artificial Intelligence interface to be used to train and select whether specific images represented explicit material or not. Ten potential vendors were reviewed, and Google Automl Cloud® was selected for training and verification testing. Unfortunately, it was difficult to obtain a sizable enough archive of approved images to complete the originally envisioned training and testing program. A modest-sized image data base was finally secured, and the code was successfully tested with a small data set, even though the results did not contain enough samples to establish the commercial-level reliability required for further testing.

Keywords

artificial intelligence (AI); censoring images; pornography; selection effectiveness; training; youth web viewing

Introduction

This paper presents the author's experiences during a self-directed research program through a Rising Scholars Multidisciplinary Engineering Research Fellowship (MERF) project under the supervision of Dr. Robert Stwalley of Agricultural & Biological Engineering to create an artificial intelligent program to prevent the viewing of pornographic images by youth. In this day and age, technology is a societal tool that can work wonders for everyone, but it is a necessary evil that can also work against us. The growing dependency of society on technology has started to have an adverse effect on our children. Children are constantly using technology, and parents rarely know what they are using it for. As a result, parents are generally left in the dark about their children's internet viewing habits, while their kids may be visiting some very questionable sites on the web. Most parents want to keep their children safe, and therefore they would like to know how to keep them from going to questionable sites, even when they are unable to actively supervise their web browsing. The researcher has been looking for new ways to work on

electronic security and may have found the solution for this particular issue: Artificial Intelligence (AI). Many companies have started to use AI to block-out 'not safe for work' (NSFW) images in order to keep their commercial work environment, professional and coeducationally-friendly office space. Unfortunately, it may be much harder than anticipated to employ AI in the near term for various reasons. In research done by Stephen (2019), he stated that training an AI to filter adult content is like showing a baby a ton of porn. Once you feed it the training data, it will start to learn new things about the data. The only problem is it could start flagging content that is not adult content. Therefore, different neural network AIs must be reviewed to see which type of code would be the most suitable for a web browser extension reviewing images. The ease of download and use would be vital features for the general acceptance of any screening code for parents who want to feel safer about their children being online.

There are a few platforms that have been trying to implement and train certain AIs in order to block-out explicit content from their apps. Software is being tested to lower the amount of unwanted explicit images that are sent to users. Khalid (2019) decided to start work on AI-based filters, because a study done by the dating site, Bumble®, saw that '57% of women felt sexually harassed on most dating apps,' and he predicted, 'Artificial intelligence will soon weed-out any NSFW photos that a Bumble® match sends to you. Bumble® intends to track-down offending users and block them on their app permanently, levying a potential fine of $500 (Khalid, 2019).

This research extends upon other AI and explicit image studies that were previously done, but more specifically, it looked into which neural networks would work best for a particular web browser extension. This will be a quantitative study to see which neural network code would be a better fit for this specific use. This research will answer the following primary research questions:

1. How will the software work in the form of a web browser extension?

2. Are there some specific neural networks that work way better than others?

3. How can the AI be made more effective at detecting mature content?

In order to achieve the most effective results, a large group of new users was originally planned to verify the AI program's training of the web browser extension in a home computer environment. Unfortunately, as will later be described, the training process was more extensive and less successful than anticipated.

Methods

Articles in the Literature

The researcher examined multiple articles in order to become well-versed with AI applications in cybersecurity (Forsey, 2018; Khalid, 2019;

Shaleynikov, 2018; Stephen, 2019; Twenty-five Clients, 2019). These articles explained in detail how various AI codes were implemented into their apps and how successful they were in their generalized training. These articles provided insight as to whether a particular AI was appropriate for the current planned use and whether it could be modified to fit the defined needs. These articles explained how the multiple different neural network Application Programming Interfaces (API) worked. This information provided a comparison of the various codes and a determination as to which ones would better suit this particular project. Five different articles were reviewed that compared over 20 neural network APIs.

The Tool

The API selected was the Google Cloud Vision® API. Ease of use was the primary selection criteria. This code was the simplest to train. All that needed to be done was to find the image to be blocked, input multiple images of that item, and then train the machine to determine whether the image is the prohibited item or not.

Curation of images

In order to find a large quantity of one type of images, a mass photo downloader was used to grab all of the matching specific search images that Google® had to offer. After this step was completed, the next task was to repeat this process for images that were not in the original set of images. Once these images were collected, they were sifted-through to make sure all of the images were, in fact, usable for training and testing. Once the images were fully organized, identified groups were created. These were labeled as 'hot dog' and 'not hot dog'. At this point, the AI could begin "training" on the data set. Once 'trained', The AI can be tested in use to see if it properly classifies new images into the two selected categories.

Validation Test

Once the program has been validated by the researcher, it was planned to test it with a group of people and see how well it works. This would have allowed an out-of-house review of the AI training, to see if there were any modifications that should be made to the program before it was more widely distributed for further testing. At least 20 different participants ranging in age and web usage were targeted to be selected for testing assistance. After a week of testing, a survey was to be sent the users to determine if the product was ready to go to market, as well as whether there were internal flaws that needed to be examined and corrected. Unfortunately, limited training sample sizes for the AI and the failure of properly training the code prevented the work from moving into this more robust stage of testing.

Data Analysis/Results

How will the AI work in the form of a web browser extension?

The researcher looked at many neural network API's before finally deciding on the one used to try the actual

discrimination application. There were many pros and cons to every API examined, and correspondence between the projected software mission and the various different codes' features were noted. The identified selection characteristics were: if the API was open source, if it was easy to use, if it was fast, and its ease of implementation.

Are there some neural networks that work way better than others?

A quantitative weighted-decision matrix comparison analysis was conducted on which neural network API would be used for this project. There were ten competitors that seemed to be potentially useful in this case. Google's Automatic Cloud Vision® was a good candidate by the sheer fact that it would be very simple to implement into an existing browser as a Google® extension, allowing for the encryption of passwords and other important information. An unfortunate issue with Cloud Vision® was that it used its own unique programming language for the code. Users would have to learn this new higher-level language in order to implement a process with this AI. Microsoft CNTK® was also a respectable choice, with a very fast open source API. The problem with using this particular code is that it potentially will not translate as well into a Google Chrome® browser extension. With CNTK®, a new programming language must also be learned in order to use it. This will likely be a bit of a challenge for those who cannot pick-up a new computer language quickly. It was finally determined that the evaluation factors for the APIs should be open source status, ease of use, speed of operation, and the effectiveness of the API at the current sorting implementation, and they were all given respective weights of 1.00. Other weighting choices in the analysis could easily be implemented, if experience dictated that one particular factor was more or less important than the others. Table 1 presents the weighted quantitative evaluations of the various APIs.

Table 1 - Quantitative analysis of various commercially available application programming interface (API) software.

API	Open Source	Ease of use	Quickness	Implementation	Total
Tensorflow®	✖	✔	✖	✖	1
Microsoft CNTK®	✔	✔	✔	✔	4
Caffe®	✔	✔	✔	✖	3
Torch®	✔	✖	✔	✖	2

Accord.NET®	✖	✖	✖	✖	0
ML pack®	✖	✔	✖	✖	1
Pytorch®	✖	✔	✔	✖	2
MxNet®	✔	✔	✔	✖	3
Chainer®	✖	✖	✔	✖	1
Auto ML Cloud Vision®	✔	✔	✔	✔	4

How can the AI be made more effective at detecting mature content?

The training to make the AI more effective at detecting objectionable images involved using a large training data set to make selections more accurate. Thousands of legally sanctioned training data images are needed to be assembled to create an accurately trained program. Not only would the training data need to be for explicit images, but the program additionally needs to not flag non-explicit images. Images that may resemble mature representations, but are actually normal objects from everyday life, must not be erroneously flagged as erotic, otherwise the credibility of the software process will be lacking. Initial training for the AI proved extremely difficult, as finding a large selection of appropriate training images was far more of a struggle than anticipated. A more modest-sized sample set was finally acquired, and this allowed a minimal proof-of-concept evaluation on a small representative data set to be completed.

Will the AI be effective at all?

Google's® neural network API was selected to demonstrate its accuracy after training. These results are from testing the neural network with a curated set of images determining how effective the AI truly was at image discrimination. The AI showed an outstanding 100% accuracy after training, but there were a few questions that remained following the tests. The results may have been perfect, but the low number of training photos limited the projected reliability of the AI. The confidence report output for the AI on this limited data set is shown in Figure 1. This result is likely skewed by the low sample size. It was difficult to find more quality training photos, and the effort required to secure this piece of the project was clearly underestimated.

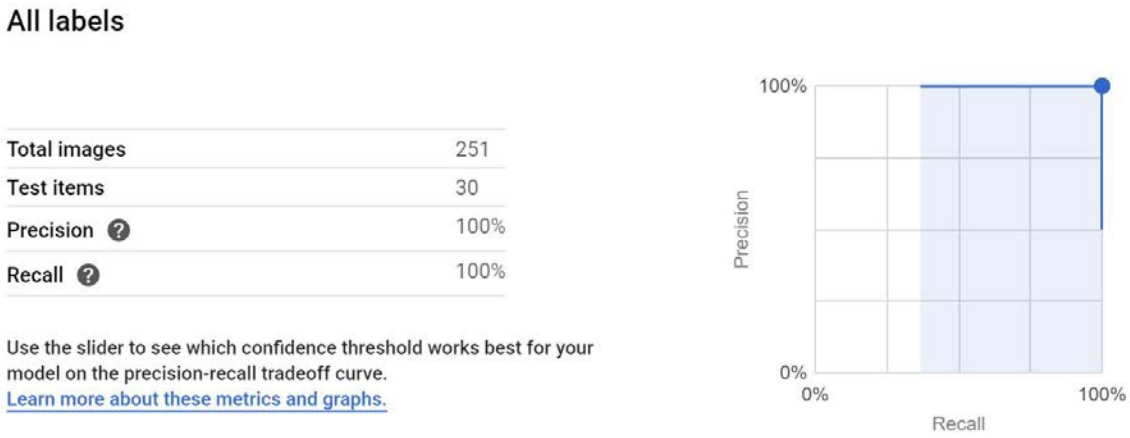

Figure 1 - The precision of the current research.

After the training had been completed, one image from the internet was fed into the program and run. The AI clearly was able to detect what type of photo it was accurately and quickly. However, the AI did not have sufficient training to be able to establish a commercial-level software confidence for further testing. The code seemed to have the potential to have been more effective, if it could have been trained with a more extensive database. Although the developed app did not ultimately progress into a distributed system test for blocking adult images, it was still an acceptable proof-of-concept software demonstration, which was effective at detecting a new image to be blocked within a personal computer environment based on its previous training.

Conclusion/Recommendation

After analyzing the literature and the data on the different types of neural network APIs that are available, it was reasonably simple to thin the prospective field with a weighted decision matrix analysis. The two best options for a web browser extension would be either Google's Automl Cloud Vision® or Microsoft's CNTK® neural network. The image recognition software in these two programs is amongst the best available in the current commercial marketplace. Google's® neural network was chosen, because it was the easiest to use and implement as a web browser extension. After working with the Google® API, it was determined that this was, in fact, probably the best choice and possibly the most accurate.

The researcher would like to see an extended, independent study on which API would be better in this application, but believes that one of the best was tested. Google Automl Cloud Vision® was simple to use and is efficient in execution. Although the other APIs may do similar things, this one was superior, because of its well-designed interface. It

is the new Google Chrome® interface, so it will be something that many people will be familiar with, as Google® is the world's leading search engine (Forsey, 2018). Most people with an internet connection will find this a convenient option to install and quickly learn how to use it.

References

Forsey, Caroline. (2018) “The top 7 search engines ranked by popularity”, HubSpot, Retrieved 1 July 2021 from https://blog.hubspot.com/marketing/top-search-engines.

Khalid, Amrita. (2019) "Bumble Will Use AI to Detect Unwanted Nudes." Engadget, Retrieved 24 April 2019 from www.engadget.com/2019/04/24/bumble-will-use-ai-to-detect-lewd-nsfy-images/.

Shaleynikov, Anton. (2018) "10 Best Frameworks and Libraries for AI - DZone AI." Dzone.com, Retrieved 23 May 2019 from www.dzone.com/acticles/progressiye-tools10-best-frameworks-and-libraries.

Stephen, Bijan. (2019) "Porn: You Know It When You See It, but Can a Computer?" The Verge, Retrieved 30 Jan. 2019 from www.theverge.com/2019/1/30/18202474/tumblr-porn-ai-nudity-artificial-intelligence-machine-learning.

Twenty-five Clients. (2019) "Deep Learning Frameworks Comparison — Tensorflow, PyTorch, Keras, MXNet, The Microsoft Cognitive Toolkit, Caffe, Deeplearning4j, Chainer." Netguru Blog on Machine Learning, Retrieved 24 April 2019 from www.netguru.com/blog/deep-learning-frameworks-comparison.