Using AWS Rekognition Video Analysis to Solve a Technical Security Issue

Muhammad Maaz Khan
St. Cloud State University
St. Cloud, United States
makhan@go.stcloudstate.edu

Abdallah Abdelwahab Mohammed
St. Cloud State University
St. Cloud, United States
aomohammed1@go.stcloudstate.edu

Seeta Kumari
St. Cloud State University
St. Cloud, United States
skumari@go.stcloudstate.edu

Ikeomumu Nnenna Diana
St. Cloud State University
St. Cloud, United States
indiana@go.stcloudstate.edu

***Abstract*— Security is required in all forms of mediums, whether it be physical or digital. The key to securing something relies on having a detection system to catch all potential threats in order to ensure safety. In the era of technology we live in today, the detection process for video security analysis can be taxing and tedious. We propose a solution that requires less manual labor to go through by being more automated and reducing human efforts. Based on the research and testing using Amazon Web Services, we created an application that aids us in these efforts.**

***Keywords*— AWS Rekognition, real-time, video analysis, Amazon, software, Open AI, deep learning, automation**

I. Introduction

AWS Rekognition Video Analysis is a service available through Amazon. The initial launch of the service occurred in 2016. The service feature allows scalable computer vision to be brought to your S3 stored video, and live stream videos. Rekognition video allows an individual to "accurately detect, track, recognize, extract, and moderate thousands of objects, faces, and content from a video." [1]

The service looks at an image, using its learning-based capabilities, and can deduce what is happening in the image based off of different objects that might be in it. This particular software could be used in a variety of different situations. [1]

Deep AI (Deep Learning) is a type of machine learning based on artificial neural networks. Deep learning has been applied to fields such as; speech recognition, computer vision, language processing, bioinformatics, social networking, audio recognition, drug design, and more. Deep Ai is a type of machine learning algorithm. It uses layers, lower and higher, to extract higher level features in an image. Lower layers process edges, while higher layers identify the concept, such as letters or faces. [4]

Both of these services can be used to help combat crime, or aid decision making processes by those in governmental positions. This paper will provide some examples of how these specific services help combat criminal behavior or activity around the world and explore some of the issues that have arisen as a result of using the software in criminal justice systems or government agencies. [4]

II. Background

AWS (Amazon Web Service) Rekognition is a relatively new, as of 2016, video analysis service available to consumers on the market. The service provides users with the ability to use scalable computer video analysis. They can take any video and the service will provide them with information based on images found within the video. It

accurately captures thousands of objects, faces, and other content within a video. [1]

Some currently available features of AWS Rekognition are face analysis, emotion analysis, gender recognition, facial recognition, facial comparison, and age detection. It also offers face attributes such as smiling or non-smiling, open or closed eyes, with or without facial hair, and face position. The program is also able to recognize objects in a scene such as pets, furniture, bicycles, and thousands of other objects. [2]

III. Literature Review

Deep learning and AWS Rekognition have been used to aid decision makers, such as country leaders, with recent events like the 2019 Coronavirus (COVID-19). The 2019 Coronavirus has affected the entire planet, remote sensing data is now more important than ever to help with handling the outbreak. The European Union Commission requested that any/all satellite images that are taken in relation to the pandemic are shared for research purposes. These images give information that help people in positions of authority, make decisions that are in the best interest for those who are suffering. [3]

The images can tell scientists about traffic patterns, and if trucks carrying emergency or critical supplies will be blocked, offers knowledge about medical facility construction due to the virus,

(a) Airport before and after the COVID-19

(b) Car rental parking lot before and after the COVID-19

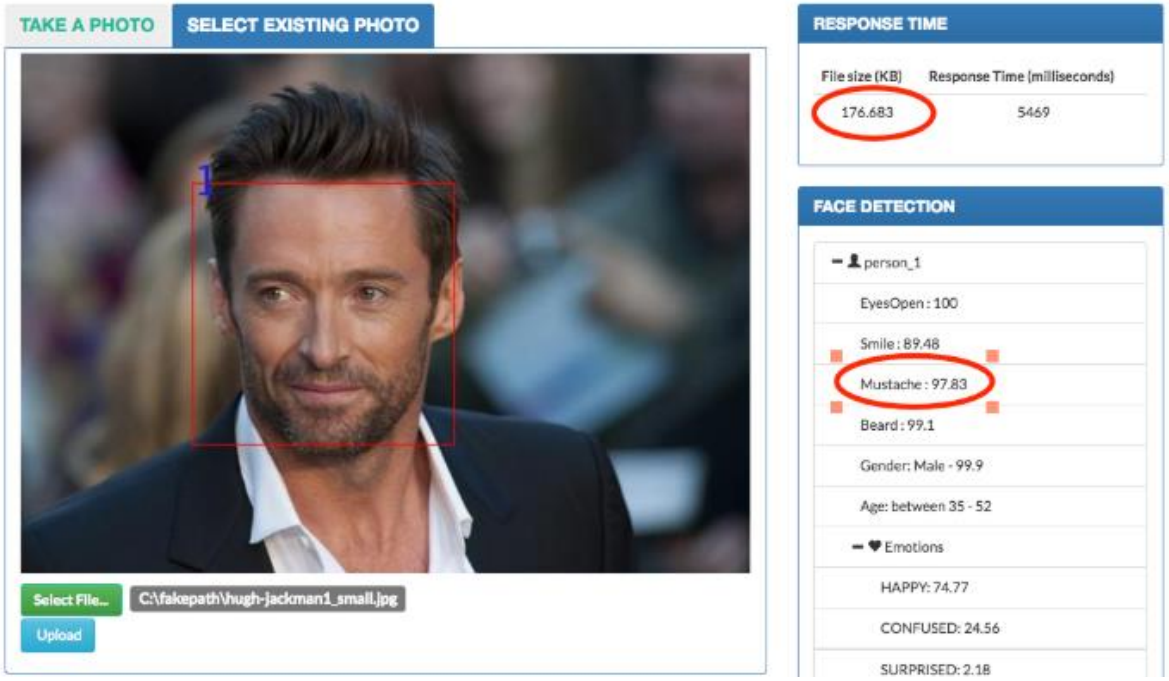
Fig. 1. COVID-19 impacts on human & economic activities. (Satellite Image 2020 Maxar Technologies). [3]

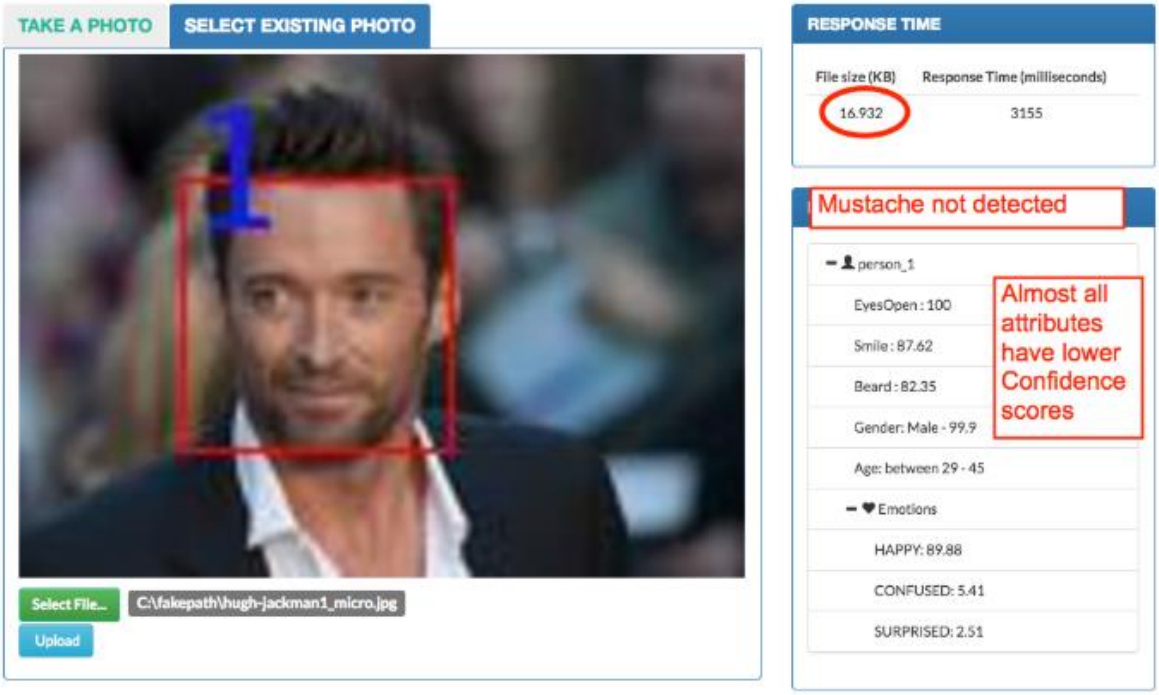
activity from different facilities and their response to COVID-19, and their resource usage, and scientists will also be able to tell if people are following the social distancing rules put in place by their government officials. They use high-resolution imagery to acquire this information. These panoramic images are taken every day, all around the globe. The image above (Figure 1) is an example of the many images that are taken every day and used to evaluate what is happening in the word due to the 2019 coronavirus, and how daily patterns are changing. [3]

In a study done by Harry Zhu and Tuan Pham when testing the software, they found that the resolution of the images was really important and changed the results from the software. With higher resolution images, AWS Rekognition was able to detect more precisely what was in the image than with the lower resolution images (Figures 2&3). [2]

Fig. 2. Celebrity, Hugh Jackman, AWS Rekognition Detection #1 [2]

Fig. 3. Celebrity, Hugh Jackman, AWS Rekognition Detection #2 [2]





A study on home security systems by using AWS Face rekognition on AWS EC2 was presented by Vyacheslav Voronov. The systems were tested successfully by using a higher accuracy rate with the help of Haar and LBP algorithms. Their results' accuracy depended on the training of faces and environmental conditions. The study described that face recognition is the most difficult topic in the field of image analysis, which has gained a great deal of attention over the last few decades due to its many applications present in various domains. [4]

Amazon Recognition from Amazon helps users to integrate deep learning-based images and video analytics into your application. This service can identify objects, individual, text, scenes and behavior as well as detect content that is inappropriate. Furthermore, Amazon Rekognition analyzes and correctly identifies faces in photographs, and company videos. This service lets you use a private collection of photos. The study previously mentioned used surveillance monitors that were placed in security labs and CCTVs were connected to the application servers. The program which processes the video streams captures, recognizes, identifies individuals, and documents employee's data in a database, and shows monitor performance. The video stream recorded will be stored on a file server. Devices and servers can be connected via WIFI or Local area network (LAN). Once the video is captured. It is then processed frame-into-frame into the application server. Once faces are detected, those are then compared with the data available in the data servers. If it matches it allows the individual in its security systems. [5]

A face recognition device was proposed based on histogram of gradient (HOG) direction that features extraction and fast principal component analysis (PCA) algorithm. This classifier study is used to extract data from the context interference, and then extract the HOG features from the image data. The PCA algorithm is used to remove the data's principal feature components. Helping vectors and not as an independent paper. In this machine algorithms were used to analyze and recognize the face. [6]

A study by Ahonen, hadid and Pietikainen, presented an effective representation of the facial image using local binary pattern (LBP) texture features. The facial image was divided into several regions from which the distributions of the LBP functions were extracted concatenated into an improved vector to be used as a facial descriptor. The invariance of gray-level monotonic adjustments and computational effectiveness made LBP a suitable for demanding tasks of image analysis. The LBP may use the texture descriptor to compose many local facial details and combine these into a global image that can be used to identify the face. [7]

A different study by Udaya Kumar presented the implementation of a low-cost wireless home protection network by using ZigBee protocol and remote Internet access. A ZigBee based star network with two nodes using the following microcontrollers: Xbee radio, ARM7, PIC and MBED was developed. With the aid of passive infrared Sensor, Gas Sensor, and Camera; an intruder motion detection, gas leakage detection and home visual surveillance were provided. The issue here is the use of many microcontrollers, the use of ZigBee based network to communicate with the base station is restricted to just 100-150 meters wide. The base station relies on the internet access only of the Ethernet. [8]

Another research study by J. SHankar Kartik, introduced two approaches, one of which is based

on GSM technology and the other is using a video web camera to detect the intruder. The first surveillance device uses a house-installed web camera that is powered by PC-installed software and uses the internet to communicate. The camera will then sense the movement of an intruder in front of the camera dimensions or given camera range. If any movement is captured the camera sends signals using internet connections to the user/ installer and at the same time it is given the ability to create a sound buzz (Sound alert system. The second security is similar to AWS because it gives a message alert. Not that AWS gives a message alert, but it identifies. Same as if the data is being stored in the data server. It alerts the system that an intruder whose data matches will be alerted via text message. In this case, it uses GSM technology for the message system. [9]

IV. Challenges

There are a lot of concerns about the software, and some articles describe that Amazon has come under fire for some of these concerns. The information available for AWS Rekognition is rather limited because the technology is still new. The reviews and more current news articles report some negative reviews about the program.

The Orlando police department in Florida, decided to try using AWS Rekognition. However, they decided to stop using it after fifteen months due to many technical issues. They originally started using Rekognition in 2017, with phase two having started in 2018, with cameras set up at their headquarters, downtown, and in one of the community recreations centers in order to track suspects in real-time. The Orlando Chief Administrative Office said that they were not able to dedicate the resources needed to the pilot that would have been needed to show any progress as a result of using the Rekognition technology, and that they have no plans to do another trial with the Rekognition technology in the future. [10]

V. Methodology

We had the option of choosing from two methods of developing the application.

The first method was, to first take a video and split it into images at an interval of every second. For example, if the video is 30 seconds long, images of every second in the video would be extracted, that would be 30 images per second. After the video had been split each image would be Analyzed using AWS Rekognition, the results would then be compiled into one output and transcribed for the user. The issue here was that the timestamps would need to be added manually based on what frame the image was from, which would be tedious to develop and unnecessary.

The second method was much less tedious, in this method the video would be analyzed as a whole using AWS Rekognition Video Analysis. The video would be uploaded to AWS S3, a storage service by amazon, the application would then take that video from S3 and analyze it. The results contain a lot of information about the video such as, the frame rate, format, duration, codec, timestamp, detected text and other information, as shown in figure X. Apart from the timestamp, the detected text and maybe, the confidence level of the detection, all other information is insignificant to the purpose of our example. The results are completely dependent on what the user wants.

```
print("Timestamp: " + str(textDetection['Timestamp']))
print("   DetectedText: " + text['DetectedText'])
print("   Type: " + text['Type'])
print("   Confidence: " +  str(text['Confidence']))
 #print("   Geometry:")
# for geometry in text['Geometry']:
#     print ("       Bounding box")
#     print ("         Top: " + str(geometry['BoundingBox']['Top']))
#     print ("         Left: " + str(geometry['BoundingBox']['Left']))
#     print ("         Width: " +  str(geometry['BoundingBox']['Width']))
#     print ("         Height: " +  str(geometry['BoundingBox']['Height']))
#     print ("       Polygon")
#     print ("         X: " + str(geometry['Polygon']['X']))
#     print ("         Y: " + str(geometry['Polygon']['Y']))
 print()
```

Fig. 4. - Various information provided from the results of the analysis (Grey).

The second methodology was chosen to build the application. There were many deciding factors that led to this choice, but the most important factor was the main purpose of this project, reducing human error and human effort so systems are more accurate and less taxing. Using this method, not only saves development hours but also makes it easy for end users to understand the application.

VI. Evaluation

In order to provide the most accurate evaluation. The tool was evaluated using two approaches.

1. Using an already existing Traffic surveillance footage from YouTube, as shown in Appendix B. This would test the accuracy of the analysis obtained from the application.

2. Using a video made by an ordinary phone camera. This would test the reliability of the analysis obtained from the application.

In evaluation method 1 the video was 30 seconds long and showed multiple cars passing by on a street, the license plates could be read clearly. This video was uploaded to AWS S3, an online storage service. The application extracted the video and analyzed it to detect text, which would be the license plate numbers, the results obtained were very accurate with almost all of the detections being over 90% accurate.

In evaluation Method 2 the application was tested for reliability. A video was taken using a personal smartphone while casually walking. To get accurate test results and remove any biases, the content of the video was kept the same as the video used in evaluation 1. The video taken for this evaluation consisted of cars moving while the person filmed them from across the street. The results obtained from the analysis of this evaluation were not that great but were still feasible. The text was detected but not with good accuracy, the results were still good enough to work with.

The combined results from both of these evaluations conclude that the application is highly accurate provided that the quality of the video or image is good.

VII. Results

The results obtained were very promising, the application was successfully able to analyze a 30 second video, detect any text and transcribe that detected text, along with the timestamp the text was detected at. Allowing the user to easily search for the license plate and find the time that license plate was detected.

VIII. Conclusion

Amazon Rekognition claims that it is easy to add image and video analysis to your applications using proven, highly scalable, deep learning technology. This claim by Amazon, Inc. is accurate after using the Rekognition Python SDK. The project team was successful in recreating the tool and specifying which API and feature(s) to be used and implemented in the final demo build.

For future plans, the team would like to carry the project further and make the tool fully functional through the created Web App (Appendix A). Including, implementing a Browser-Based Upload using HTTPS POST. [12]

IX. References

[1] Walker, T. (2017 29). Welcoming Amazon Rekognition Video: Deep-Learning Based Video Recognition. Retrieved May 4, 2020, from https://aws.amazon.com/blogs/aws/launch-welcoming-amazon-rekognition-video-service/

[2] Zhu, H., & Pham, T. (2017, March 28). Leveraging AWS Rekognition for near real-time face recognition. Retrieved May 6, 2020, from https://dius.com.au/2017/03/28/leveraging-aws-rekognition-for-near-real-time-face-recognition/

[3] Minetto , R., Segundo, M. P., Rotich, G., & Sarkar, S. (2020, April 16). Measuring Human and Economic Activity from Satellite Imagery to Support City-Scale Decision-Making during COVID-19 Pandemic. Retrieved May 4, 2020, from https://deepai.org/publication/measuring-human-and-economic-activity-from-satellite-imagery-to-support-city-scale-decision-making-during-covid-19-pandemic

[4] Chippa, S. V., & Dube, R. R. (2019, August 26). AWS EC2 based Home Security System using Face Recognition.

[5] Vyacheslav Voronov, Vladimir Strelnikov, Liliya Voronova, Artyom Trunov, Andrey Vovik, (2019). Faces 2D-Recognition and Identification Using the HOG Descriptors Method. Moscow Technical University of Communications and Informatics, Moscow, Russia.

[6] Li, X.-Y., & Lin, Z.-X. (2017). Face Recognition Based on HOG and Fast PCA Algorithm. Proceedings of the Fourth Euro-China Conference on Intelligent Data Analysis and Applications Advances in Intelligent Systems and Computing, 10–21. doi: 10.1007/978-3-319-68527-4_2

[7] Ahonen, T., Hadid, A., & Pietikainen, M. (Dec, 2006). Face Description with Local Binary Patterns: Application to Face Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[8] Udaya Kumar, D. S. (march 2013). “Implementation of Low Cost Ethernet Based Home Security Using. Journal published at Algorithms.

[9] J. Shankar Kartik, K. Ram Kumar and V.S. Srimadhavan3, (2013) “SMS Alert and Embedded Network Video Monitoring Terminal”, International Journal of Security, Privacy and Trust Management (IJSPTM), Volume 2.

[10] Roulette, J. (2020, May 4). Orlando cancels Amazon Rekognition program, capping 15 months of glitches and controversy: Orlando Weekly exclusive. Retrieved from https://www.orlandoweekly.com/Blogs/archives/2019/07/18/orlando-cancels-amazon-rekognition-capping-15-months-of-glitches-and-controversy

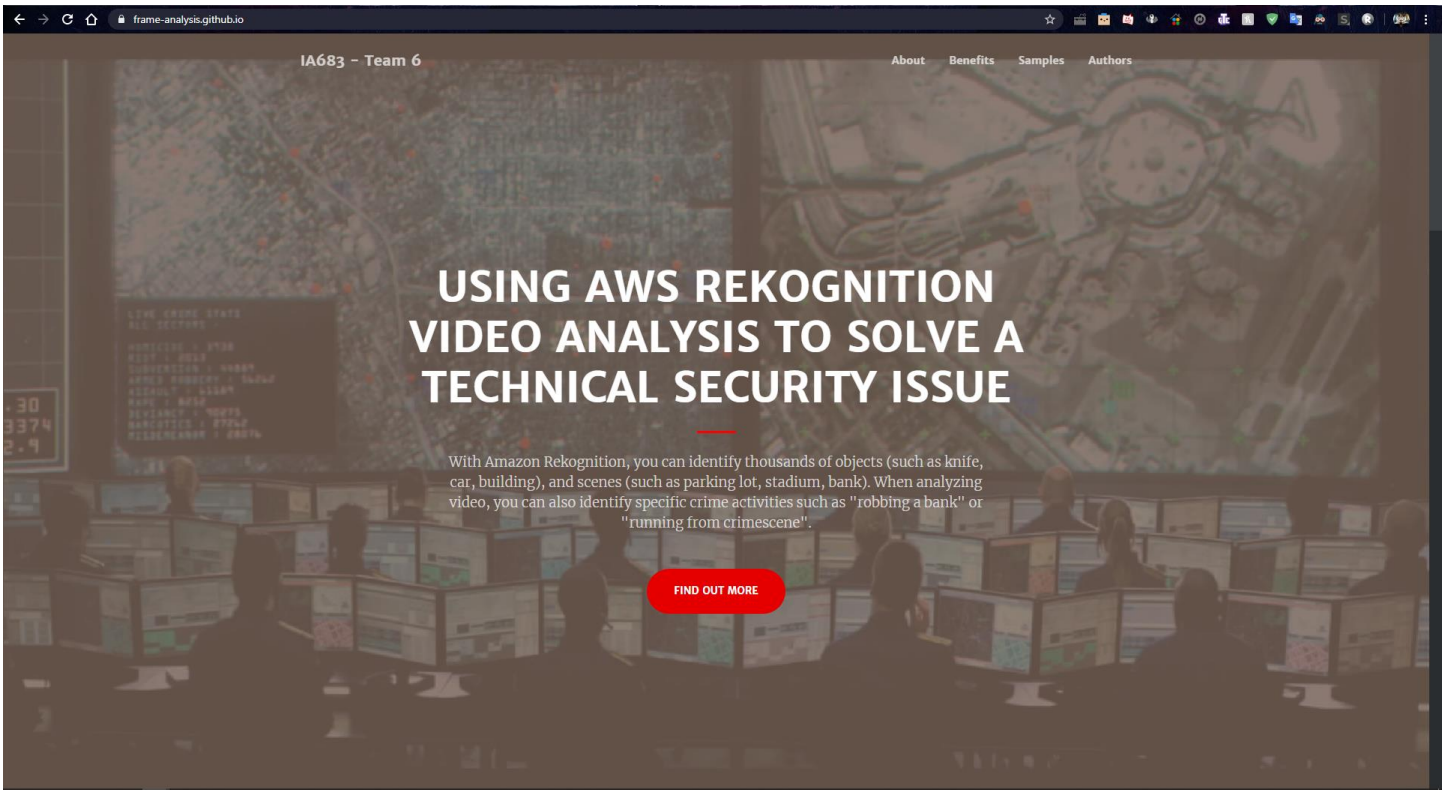
[11] Mohammed, A., & Maaz, M. (2020, May 8). Frame Analysis Demo. Retrieved from https://frame-analysis.github.io/

[12] Amazon Web Services, Inc. or its affiliates. (2020) Browser-Based Upload using HTTP POST (Using AWS Signature Version 4). Retrieved May 7, 2020, from https://docs.aws.amazon.com/AmazonS3/latest/API/s3-api.pdf#sigv4-post-example

[13] Tech Channel00001. Youtube. Praffic IP Camera video. Retrieved April 20, 2020, from https://www.youtube.com/watch?v=Gr0HpDM8Ki8&fbcl=

Appendix A

The Front-End Web App Mockup. (2020). [11]



IA683 - Team 6

About Benefits Samples Authors

Start scanning!

Start scanning frame by frame, per millisecond, by uploading a video file.

GET STARTED!

Benefits

Time Efficient

Saves time with manual labour to scan frames by becoming more autonomous.

Up to Date

All AWS dependencies are kept current to keep things fresh.

Ready to Use

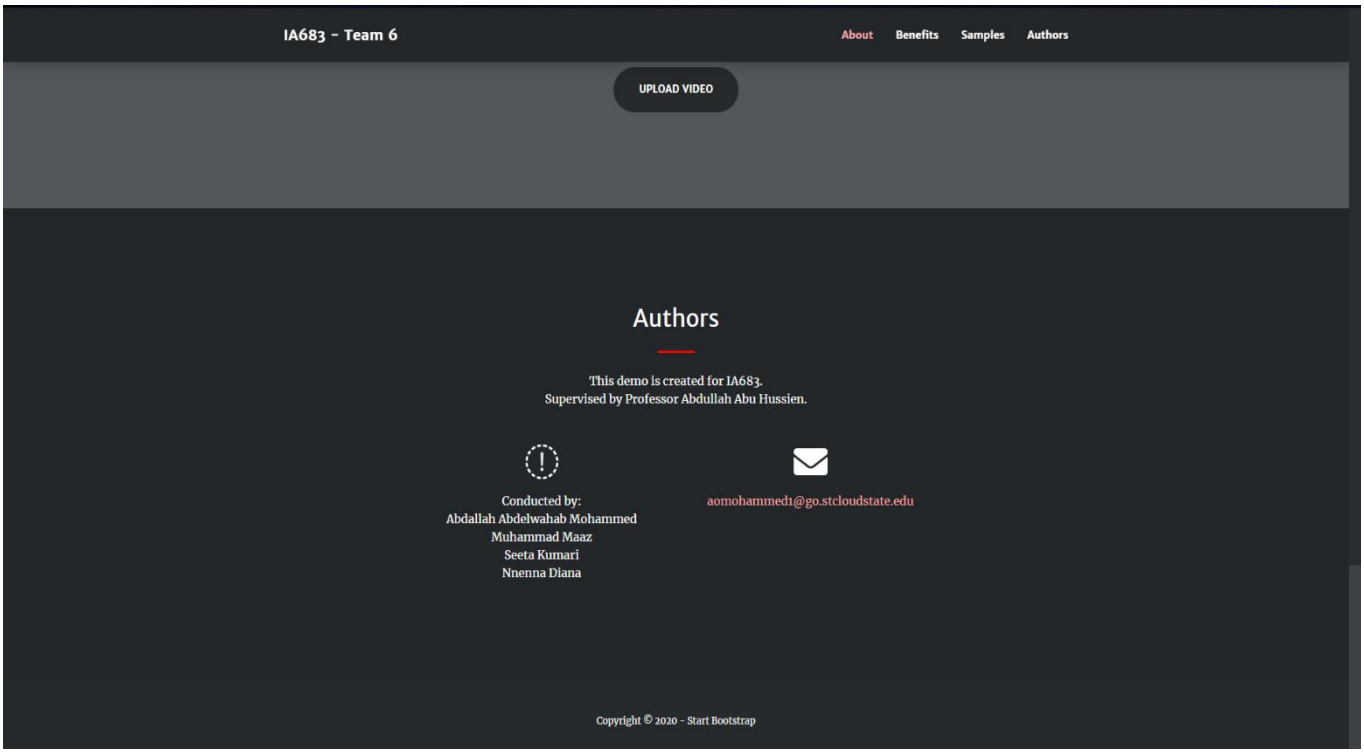
The SDK and APIs are setup to your scanning needs.

Made to Protect

The purpose of this project is to protect the public from crimes and enforce counter-terrorism.

IA683 - Team 6

About Benefits Samples Authors



Appendix B

YouTube Traffic Surveillance video. (2012). [13]