Real-Time Classification and Object Detection to provide Mass Shootings a faster response

Alejandra Vidal, Master Student in Computer Engineering
Polytechnic University of Puerto Rico
vidal4275@gmail.com
Roberto Vazquez Ferrer, Doctoral Student in Engineering and Applied Science
Polytechnic University of Puerto Rico
rjvf.007@gmail.com
Alfredo Cruz. PhD
Polytechnic University of Puerto Rico
alcruz@pupr.edu

Abstract

Nowadays, the number of public mass shootings has increased substantially. Unfortunately, with the recent high-profile active shootings, firearms safety and the response have become a pressing and severe public safety issue. This paper presents a real-time video active-shooting detection solution that could implement for police, military, government, schools, homeowners, shopping centers, and more. The system aims to classify, detect, and alert for a faster and better response to possible crimes and mass shootings. The designed application incorporates a simplified machine learning/artificial intelligence approach creating models with TensorFlow Lite to classify acoustics with video analysis using object detection. The application was tested on Android mobile phones, obtaining an accuracy of 89.52% with an average response time for Active shooting of 3.895 seconds.

Keywords: Active shooting, Acoustic Classification, Object Detection, Machine Learning, TensorFlow Lite.

Introduction

The FBI defines an active shooter as an individual actively killing or attempting to kill people in a populated area. Recent active shooter incidents have underscored the need for a coordinated response by law enforcement and others to save lives (FBI, 2019). In recent high-profile active shootings, firearm safety and response have unfortunately become a pressing issue and a severe subject of public safety (Vidal, Vazquez, & Cruz, 2022). The FBI designated 40 shootings in 2020 as active shooter incidents (FBI, 2021). Thirty-eight were killed and 126 wounded, excluding the 13 shooters. Active-shooter situations are becoming increasingly common in the United States. From 2019 to 2020, there were 40 active shooter incidents, with 164 casualties.

Law enforcement officers must respond to these situations quickly, resulting in fewer casualties (Stewart, 2017). It is challenging to prepare for active shooter incidents. However, with the increasing number of shootings yearly, workplaces and schools must prepare and train to react and respond to these situations. Performing a risk analysis to analyze and quantify the risk of a shooting at a specific school or workplace could be beneficial in many ways. The investigation results could benefit both civilians and law enforcement officers in their training, along with giving information to the administrators or leadership at the school or workplace to help them make better decisions regarding risk mitigation strategies (Stewart, 2017). Figure 1 shows the alarming and increasing number of mass assassinations in the United States since 2017.

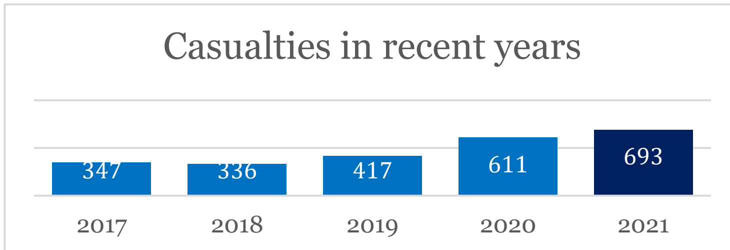

Figure 1: Chart of the mass shooting in recent years

It is difficult for law enforcement agencies to determine the exact crime location and even more difficult to respond quickly. Gunshot analysis has received significant attention from both the military and scientific communities. Acoustic analysis of gunshots can provide helpful information, such as the shooter's position, the projectile trajectory, the caliber of the gun, and the gun model (Raponi, Oligeri, & Ali, 2022). Every second after a crime continues, more lives are at stake. The purpose is to use technology to make the world safer and make protection benefits accessible to all. Artificial Intelligence and machine learning using mobile sensor technology make detection more affordable and replicable.

This project intended to reuse or phone for society's protection. Easy to keep connected to mass usage, social environment, and more. This paper presents the electronic devices model to recognize active shooter signatures to protect against hostile actions based on phone microphones and images.

Related Works

There have been several efforts to create efficient and accurate weapon detection systems to provide a faster response. Existing systems are expensive and need to be escalated. Most of them are used by the military and private sectors at an expensive cost. The existing solution requires the installation of equipment and sensor in multiple areas. A few examples of the most used shotting detection system did found.

A. Guardian Indoor Active Shooter Detection System

Guardian combines acoustic gunshot detection technology with infrared sensor muzzle flash detection to provide immediate detection and alert with zero false alerts. Integrating Security Center and Guardian Indoor Gunshot Detection enables customers to immediately react to a gunshot event and activate their emergency response. Guardian Gunshot Detection leverages audio and infrared (IR) detectors to immediately identify a gunshot event and its location within Security Center. Create unique custom events and trigger immediate distress signals and emergency responses: Defense Advanced Research Projects Agency (DARPA) and a principal defense contractor. The technology has already saved lives on battlefields worldwide and is now being implemented by businesses worldwide to protect their people and assets (Genetec, 2022).

B. 3xLOGIC's Gunshot Detection System

3xLOGIC Gunshot Detection identifies a gunshot using concussive force recognition technology; the system is a self-contained device. Utilizing a range of 75 ft. radius with a 360° coverage. The solution minimizes the number of sensors needed to be affordably covered. Gunshot detection does not utilize microphones, infrared sensors, or complex analytics - just an affordable sensor solution that is part of active shooter safety and security measures. The Gunshot Detection's infrastructure does not need software, servers, or a relay board (3xLogic, 2022).

C. ShotSpotter

The ShotSpotter mass shooting detection system can be costly, ranging from $65,000 to $85,000 per square mile, with a minimum coverage area of three-square miles. Equipment and data are available only via lease. Leasing equipment can reduce upfront costs, which helps make the price more manageable for some departments. ShotSpotter owns the data collected with its equipment and leases access to the data to police departments with an annual subscription. These agreements may prohibit departments from releasing the data so the public is unaware of incidents captured by ShotSpotter or subsequent arrests. In addition, ShotSpotter leases are structured so that police departments lose data access upon contact termination; this can decrease competition in the market (Gatens & Reichert, 2019).

Summary: Guardian Indoor Active Shooter Detection System, 3xLOGIC's Gunshot Detection System, and ShotSpotter are all designed to detect and alert authorities to the sound of gunshots in real-time. While these systems use advanced technology, such as acoustic, they do not utilize artificial intelligence (AI). Artificial intelligence can analyze multiple data sources and provide insights into potential gun violence threats before they occur. Also, this can include analyzing social media posts, online forums, and other online activity to identify potential threats and analyzing video surveillance footage and other data sources like acoustic to provide more comprehensive situational awareness.

Methodology

We are combining the strengths of both acoustic and image-based active shooter detection techniques and using deep learning to learn discriminative features for identifying an active shooter in a given environment.

1. Data Collection: Collect an extensive audio and image recordings dataset that simulates an active shooter scenario in environments like the City of Houston (2012), Muscogee County School Training (2015), Surviving LA County Sheriff (2015), OSU Surviving an Active Aggressor (2018), GSU Run Active Shooter (2019). The audio recordings should capture gunshots and other relevant sounds, while the image recordings should capture the shooter and potential victims.
2. Data Preprocessing: Preprocess the audio and image recordings to prepare them for training the deep learning model.
3. Audio Processing Model: Use a Convolutional Neural Network (CNN) like YAMNet to process the audio recordings and detect gunshot sounds. The system does create from the previous work from Vidal, A., Vazquez, R., & Cruz, A. (2022).
4. Image Processing Model: Use a separate CNN to process the image recordings and detect the presence of a shooter. This model could be trained using transfer learning, where a pre-trained model such as AutoML, MobileNet-v2, Yolov4, and Yolov5 fine-tuned on the active shooter dataset.
5. Fusion: Combine the output of the audio and image models using a fusion method such as early, late, or mid-level fusion. The fused output should indicate the probability of an active shooter in the environment.
6. Evaluation: Evaluate the performance of the combined model on a held-out test set using standard metrics such as precision, recall, and F1 score. Fine-tune the model based on the evaluation results.

Proposed System Description

The designed application partially implements the system presented, adding real-time features for audio classification and object detection. This prototype uses TensorFlow Lite audio classification, object detection combined with GPS location, provided by the mobile phone embedded GPS. The system's architecture consists of three main modules: the Audio Classification module, the Object Detection module, and the locate alert module. Altogether, these three modules are executed periodically in a sequential manner on the mobile device.

A. Audio Classification Module

The shooting detection algorithm does execute every 1000 milliseconds over self-contained audio data segments. This splitting approach simplifies the complete trajectory and resets the cumulative errors by considering independent traces. This value offers a good trade-off between user feedback and a high processing load. The application uses TensorFlow Lite audio classification with YAMNet. TensorFlow is a machine-learning library created by Google to allow for easier machine-learning model creation, training, and deployment. TensorFlow Lite is part of TensorFlow, created to run inference on phones or embedded devices. TensorFlow Lite converts a TensorFlow model into TensorFlow Lite models, which are smaller and able to run more efficiently. Optimizing the model is also easy with TensorFlow Lite as it offers post-training quantization. TensorFlow Lite does not require the entire TensorFlow library to be imported and can utilize the TensorFlow Lite interpreter package, which essential core features run inference on Lite models.

B. Object Detection Module

The initial module in the architecture oversees gathering the audio from the input video prerecorded for this experiment. If an alarming sound is detected, the object detection module is activated. The object detection module can do activated by default without the audio classification system. The object detection retrieves the images from the video, obtaining and resizing to Bitmaps of 320 x 320. It extracts images every 1000 milliseconds to be processed through the object detection module, as explained in algorithm 2.

Algorithm 2. The object detection algorithm consists of image extraction from the video; audio is intended to activate the video object detection module if necessary.

```
procedure OBJECT DETECTION
  Initialize & Start sampling timer (1000ms)
  // Extract bitmaps images from the video every 1000ms
  VideoData []  ← Video Input
  Initialize & Start sampling timer (1000ms)
      function onTick()
          // Interval analysis Image Extraction
          Events [ ] ← Object Detection (VideoData [ ])
          Filter Results [ ] by threshold
          GPS request
          Locate and Alert if founded
          Clear VideoData [ ] and Events[ ]
          Re-start Countdown timer
      end function
  end procedure
```

The Object Detection Module is the additional module of the system. Also, the real-time shooting detection algorithm is based on the model shown in Algorithm 2. It consists of two sub-algorithms: object detection from image extraction of real-time video processing. Object detection is also a computer vision technique in which a software system can detect, locate, and trace an object from a given image or video. The particular attribute regarding object detection is that it identifies the class of the object and its location-specific coordinates in the given image. The location is pointed out by drawing a bounding box around the object. The bounding box may or may not accurately locate the object's position. The ability to locate the object inside an image defines the performance of the algorithm used for detection. Object detection using TensorFlow, the feature extraction is carried out for each segmented rectangular area to predict whether the rectangle contains a valid object (Great Learning Team, 2022). This module extracts images from the video inputs and analyses them to detect weapons from the bitmap's images.

Automated machine learning (AutoML) is the process of automating the tasks of applying machine learning to real-world problems. AutoML potentially includes every stage, from beginning with a raw dataset to building a machine learning model ready for deployment. AutoML was proposed as an artificial intelligence-based solution to the growing challenge of applying machine learning (Wu & Wu, 2022).

Figures 5A-5F show the scenario that tested the objected detection module. First, Figure 5A shows that the object-detection model detected a weapon with 84.77% from Surviving an Active Shooter for the City of Houston (2012). Second, Figure 5B presents the Weapon Detected with a Confidence of 82.42% for Muscogee County School Training (2015). Third, Figure 5C presents the Weapon Detected with a Confidence of 93.62% for Surviving LA County Sheriff (2015). Fourth, Figure 5D presents a Weapon Detected with a Confidence of 85.94% for GSU Run Active Shooter (2019). Figure 5E presents the Weapon Detected with a Confidence of 94.28% for OSU Surviving an Active Aggressor (2018). Finally, Figure 5F presents Weapon Detected with a Confidence of 87.53% for ALICE Training at Ohio University (2013) using the model detection trained with MobileNetv2.



Fig. 5A Fig. 5B Fig. 5C


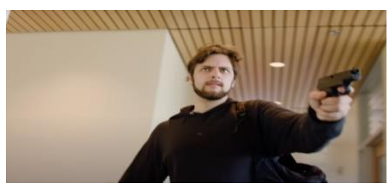
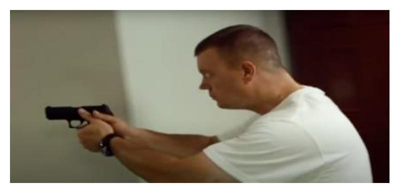

Fig. 5D Fig. 5E Fig. 5F

Figures 5A-5F: Weapons Detected with the application using MobileNetv2.

Google Cloud AutoML is a collection of machine learning that enable developers with limited machine learning expertise to train models. AutoML helps developers build custom image recognition models, including those without machine learning expertise. The AutoML features training of custom vision models with a simple graphical user interface and state-of-the-art performance that allows users to generate custom machine learning models in short minutes. In addition, the system adds human labeling by supporting a team of in-house human labelers that will examine images' custom directives and classify them, thus keeping them private. Also, Users can leverage AutoML and transfer learning technology to produce high-quality machine-learning models (Gomila et al., 2018).

The model was trained to utilize AutoML Vision. The AutoML Vision also allows training machine learning models to classify images according to the definition labels and deploy low-latency, high-accuracy models optimized for edge devices. The weapon model does train from labeled images and evaluates performance. The trained model can export as a Tensorflow Lite model to be used in the Android version.

The models were retrained using MobileNet-v2, Yolov4, and Yolov5. The reason is that MobileNetV2 is a convolutional neural network architecture that seeks to perform well on mobile devices. It is based on an inverted residual structure where the residual connections are between the bottleneck layers. The intermediate expansion layer uses lightweight depthwise convolutions to filter features as a source of non-linearity. The architecture of MobileNetV2 includes a convolution layer with 32 filters followed by 19 residual bottleneck layers (Sandler et al., 2018).

YOLO is one of the most famous object detection algorithms due to its speed and accuracy. The second model trained was using YOLOv4. Also, YOLOv4 prioritizes real-time detection and conducts training on a single GPU. The authors intend for vision engineers and developers to easily use their YOLOv4 framework in custom domains. All the YOLO models are object detection models trained to look at an image and explore a group of object classes; when the object is discovered, the system encloses them in a bounding box with its class (Solawetz, 2020). The third model used YOLOv5. YOLOv5 is another member of the (YOLO) family of computer vision models. YOLOv5 is commonly used for detecting objects. Yolov5 improved its performance and architecture with a high-level Object detection architecture (Gutta,2021).

C. Location and Alert Module

This module performs when the object detection step detects a possible shooting or mass shooting. The location module manages the location and data class being a geographic location with latitude, longitude, timestamp, and other information such as bearing, velocity, and altitude. All areas generated through the location manager are guaranteed valid latitude, longitude, and timestamps. The Kalman filter improves location accuracy by combining the network and GPS providers. In addition, an inform will be sent to pre-determined users in real time with the discovery and results of the module.

III. MODELS, FUSIONS, and EVALUATION

The system was tested using several active shooting training videos. The first video mass shooting sample was Surviving an Active Shooter (City of Houston, 2012). The second video was Muscogee County School Training (Muscogee County School Training, 2015); this video is a short narrative that depicts a High School shooting situation and examples of proper and improper responses. It is part of a training program for the Muscogee County School District in Columbus, GA. The third video was Surviving an active shooter, LA county sheriff (LA County Sheriff, 2015). Los Angeles County Sheriff’s Department created a video to help people understand what to do if they do attack by a shooter at school, work, or in public. CSUMB Active Shooter Training (California State University, Monterey Bay, 2016) presented training for students and members for an active shooting scenario.

The video of GSU Run Active Shooter (Georgia Southern University, 2019). Georgia Southern University Police Department’s Chief of Police Laura McCullough goes over the three strategies anyone on campus should use if such a situation ever takes place: run, hide, or fight. In the video of OSU Surviving an Active Aggressor (Ohio State University, 2018), the university presents how an active aggressor could strike any place at any time. Ohio State’s Department of Public Safety shares the “Surviving an Active Aggressor” video to educate the campus community. The 9-minute video updates the university’s original safety tutorial, released in 2015.

The last video was presented by the Ohio University Police Department and ALICE, providing students and faculty with the information and tools they need if an active shooter attempts to impact school safety. This video gives a briefing summary of the ALICE Training method. Evacuate and reach a safe distance, then call 911. Active shooter situations often require lockdown. Sharing tips about effective lockdown solutions from the experts at ALICE Training, then learning the steps to evacuate safely and how to defend against an armed intruder ALICE Training at Ohio University (Ohio University Police Department, 2013).

The video presented in Figure 6 is a short narrative that depicts a High School shooting situation and examples of proper and improper responses. It is part of a training program for the Muscogee County School District in Columbus, GA. The MobileNetv2 model successfully detected it with a confidence of 92.97%. The first image captures the video, and the second screenshot represents the application. In the application, the image is resized to 224 x 224 to be analyzed by object detection through TensorFlow Lite.

Figure 6: Active Shooting Detected with Confidence 92.97% of the Muscogee County School Training video and a screenshot of the application.

Results

For the experiments, four models were trained and analyzed. The models were trained using MobileNet-v2, Yolov4, and Yolov5. The videos used in this experiment were college's earlier and high school's active shooting training videos posted on YouTube for students' education purposes, with a perfect audio and video scenario. Table 1 shows the results of the matching percentage of the detection of the active shooting. For instance, the first test gave an accuracy of 84.77% with the model MobileNetv2, 84.98% accuracy with Yolov4, and 85.66% accuracy with the model Yolov5 during the first video Surviving an Active Shooter. The test was repeated for all the discussed models for each training video.

Table 1. Results from the object detection of the active shooting videos

Video Object Classification Results	MobileNetv2	YoloV4	YoloV5
Surviving an Active Shooter	84.77	84.98	85.66
Muscogee County School Training	92.97	90.61	93.31
Surviving LA County Sheriff	93.62	94.17	95.22
GSU Run Active Shooter	85.94	90.23	88.48
OSU Surviving an Active Aggressor	94.28	87.94	93.95
ALICE Training at Ohio University	87.53	79.65	88.02
Accuracy	**89.85**	**87.93**	**90.77**

Table 2 shows that the evaluation of the system shows a combined detection accuracy of 89.52%. MobileNetv2 with 89.85% accuracy for evaluated models, YoloV4 evaluated with 87.93% accuracy, and YoloV5 with 90.77% accuracy. The results of each type of model were shared in one table to obtain a system average. This system was tested using prerecorded training videos with a regulated time frame for better results. The best results were obtained with the model of Yolov5 having an accuracy of 90.77%.

Table 2. The average result for detecting the active shooter

Type	**Results**
MobileNetv2	89.85
YoloV4	87.93
YoloV5	90.77
Average	**89.52**

Table 3 shows the results in seconds (detection time) for the application to spot the active shooting. We chose the control time as the threshold for active shooting. The model's times are how long in seconds after the active shooting was detected. For instance, the active shooting in the video Surviving an Active Shooter occurred a time 1:16 (one minute, sixteen seconds), MobileNetv2 spent 4.931 seconds to detect the shooter, YoloV4 spent 4.641 seconds to detect the shooter, and YoloV5 spent 2.510 seconds to detect the shooting. The test was repeated for all the discussed videos for each model.

Table 3. Results response time of the active shooting videos

Video Object Classification Results	Control time	MobileNetv2 in secs	YoloV4 in secs	YoloV5 in secs
Surviving an Active Shooter	1:16	4.931	4.641	2.510
Muscogee County School Training	15:28	5.174	4.635	1.816
Surviving LA County Sheriff	1:57	6.056	4.411	3.108
GSU Run Active Shooter	0:35	1.028	3.748	3.135
OSU Surviving an Active Aggressor	5:11	4.482	5.021	4.649
ALICE Training at Ohio University	2:07	2.166	4.319	4.282
	Average	**3.973**	**4.463**	**3.250**

Table 4 shows that the system assessment shows a combined detection time average of 3.895 seconds. Average time for evaluated models; MobileNetv2 with 3.973 seconds, YoloV4 evaluated 4.463 seconds accuracy, and YoloV5 with 3.250 seconds. The results of each type of model were shared in one table to obtain a system average. This system was tested using prerecorded training videos with a regulated time frame for better results. The best results were obtained with the model of Yolov5 having a detection time of 3.250 seconds.

Table 4. The average result of time to detect the active shooting from sample videos

Type	**Results in secs**
MobileNetv2	3.973

YoloV4	4.463
YoloV5	3.250
Average	**3.895**

Conclusions

This paper presents a model to detect mass shootings in real-time using the embedded video input (prerecorded) from smartphones exclusively. The application model includes three modules to analyze the audio data, image extraction, object detection, and notify results. A series of algorithms and mechanisms related to TensorFlow Lite does implement to detect mass shootings and alerts. The evaluation of the system reveals a combined detection accuracy of 89.52% and an average response time of 3.895 seconds.

Acknowledgment

This research has been supported by, or in part by, the NCAE-C under contract award H98230-20-1-0411.

References

Alexandria Stewart. (2017). *Active shooter simulations: An agent-based model of civilian response strategy*. Library Iowa State University. Retrieved November 9, 2022, from https://dr.lib.iastate.edu/handle/20.500.12876/29804/

California State University, Monterey Bay. (2016, February 29). *CSUMB Active shooter training video*. YouTube. Retrieved November 13, 2022, from https://www.youtube.com/watch?v=0NGBTdnRdQw

City of Houston. (2012, July 23). *Run. hide. fight. surviving an active shooter event*. YouTube. Retrieved November 13, 2022, from https://www.youtube.com/watch?v=5VcSwejU2D0&t=80s

FBI. (2016, June 9). *Active shooter safety resources*. FBI. Retrieved November 9, 2022, from https://www.fbi.gov/about/partnerships/office-of-partner-engagement/active-shooter-resources

FBI. (2021, July 1). *Active shooter incidents in the United States in 2020*. FBI. Retrieved November 9, 2022, from https://www.fbi.gov/file-repository/active-shooter-incidents-in-the-us-2020-070121.pdf/view

Gatens, A., & Reichert, J. (2019, December 10). *Police technology: acoustic gunshot detection systems. Illinois criminal justice information authority*. Retrieved November 9, 2022, from https://icjia.illinois.gov/researchhub/articles/police-technology-acoustic-gunshot-detection-systems

Genetec. (2022). *Detects gunshots in under one second with zero false alerts*. Partner solutions. Retrieved November 9, 2022, from https://www.genetec.com/partners/partner-integration-hub/shooter-detection-systems/guardian-indoor-active-shooter-detection

Georgia Southern University. (2019, May 17). *Georgia Southern University Run, hide, fight active shooter training*. YouTube. Retrieved November 13, 2022, from https://www.youtube.com/watch?v=uSRP9ZsGqiA

Gomila, J., He, T., Sewell, D., Rostyslav, K., Karpinski, J., Christiansen, J., Faraguna, C., & Manipula, M. (2018). *Google AutoML, Suite of machine learning products developed by Google that enables people to train machine learning models*. Retrieved November 9, 2022, from https://golden.com/wiki/Google_AutoML-R9DGBVV

Gunshot detection. *3xLogic*. (2022). Retrieved November 9, 2022, from https://www.3xlogic.com/products/gunshot-detection

Gutta, S. (2021, August 4). *Object detection algorithm-YOLO V5 architecture*. Medium. Retrieved November 15, 2022, from https://medium.com/analytics-vidhya/object-detection-algorithm-yolo-v5-architecture-89e0a35472ef

LA County Sheriff. (2015, March 17). *Surviving an active shooter - LA County Sheriff*. YouTube. Retrieved November 15, 2022, from https://www.youtube.com/watch?v=DFQ-oxhdFjE

Muscogee County School District in Columbus, GA. (2015, July 14). *Active shooter video*. YouTube. Retrieved November 13, 2022, from https://www.youtube.com/watch?v=cO5FJ0v_tuc&t=467s

Ohio State University. (2018, August 13). *Surviving an active aggressor*. YouTube. Retrieved November 13, 2022, from https://www.youtube.com/watch?v=2zfiQAk927s

Ohio University Police Department. (2013, August 1). *Alice Training at Ohio University: Active Shooter Training*. YouTube. Retrieved November 13, 2022, from https://www.youtube.com/watch?v=5-2-Tb-wXBI

Raponi, S., Oligeri, G., & Ali, I. M. (2022, April 6). *Sound of guns: Digital forensics of gun audio samples meets artificial intelligence - multimedia tools and applications*. SpringerLink. Retrieved November 9, 2022, from https://link.springer.com/article/10.1007/s11042-022-12612-w

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L.-C. (2018). MobileNetV2: Inverted residuals and linear bottlenecks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvpr.2018.00474

Solawetz, J. (2020). Yolov4 - an explanation of how it works. Roboflow Blog. Retrieved November 15, 2022, from https://blog.roboflow.com/a-thorough-breakdown-of-yolov4/

Team, G. L. (2022, August 22). *Real-time object detection using Tensorflow*. Great Learning Blog: Free Resources what Matters to shape your Career! Retrieved November 9, 2022, from https://www.mygreatlearning.com/blog/object-detection-using-tensorflow/

Vidal, A., Vazquez, R., & Cruz, A. (2022, May). *Real-Time Classification and Object Detection to Provide Mass Shooting a Faster Response,* Security Conference Las Vegas, 2022.

Wu, C.-Y., & Wu, S. (2022). *AutoML Services*. WHEELJET. Retrieved November 9, 2022, from https://wheeljet-ai.weebly.com/services.html