CSE 473/573 - COMPUTER VISION AND IMAGE PROCESSING DETECTING UNAUTHORIZED INDIVIDUALS WITH FIREARMS PROJECT PROPOSAL

Vignesh Prakash
UBID : 50478782
prakash8@buffalo.edu
Github Link

ABSTRACT

Detection of threats possessed by unauthorized individuals carrying firearms plays a crucial role in improving public safety. To address this issue, the deployment of an autonomous system capable of alerting law enforcement officials of such threats is critical. Over the years, there are many computer vision algorithms for Object detection and with the help of those algorithms this project aims to develop a real-time firearm detection system that can identify unauthorized individuals carrying guns in the public and protected areas. The goal of this project is to enhance the security measures and enabling swift responses to potential threats to mitigate gun-related violence and protect the public.

1 Introduction

1.1 State of the Art

In Computer Vision, there are two major object detection systems, Faster RCNN (Region-Based Convolutional Neural Network) and YOLO (You Only Look Once). The Faster RCNN is a two stage system and YOLO is an one stage system, which means RCNN use region proposal sub-networks to generate proposals and YOLO generates proposals directly on the feature map.[1]

1.1.1 Is there any other solutions addressed this problem?

Yes, there is a research paper titled as "TYolov5[2]: A temporal YOLOv5 Detector based on Quasi-Recurrent Neural Networks for Real Time handgun detection in Video". This project is the first to apply QRNN to extracts spatio-temporal features for handgun detection. This model is built by using 8000 annotated images and claims to get a mAP (Mean Average Precision) of 90 at 56.8 fps. Link to Github Repository

Similarly, I have found another paper using YOLOv5 model to detects objects in the video feed and it is titled as "YouTube-GDD[1] A challenging gun detection dataset with rich contextual information". This model uses about 5000 well-chosen images, in which 16064 instances of gun and 9046 instances of person are annotated. This model claims to get a AP (Average Precision) of 52.1.

1.1.2 What am I doing differently from the current state of the art?

From the research papers, I concluded that YOLOv5 model works better at detecting objects in real-time based on the data streams. The main difference in this project would be identifying whether the person holding the firearm is an authorized person or not. I wish to train the model with specific annotations for authorized persons and detect features like the armour with police written in bold words, which can potentially distinguish them authorized persons even if they are holding firearms.

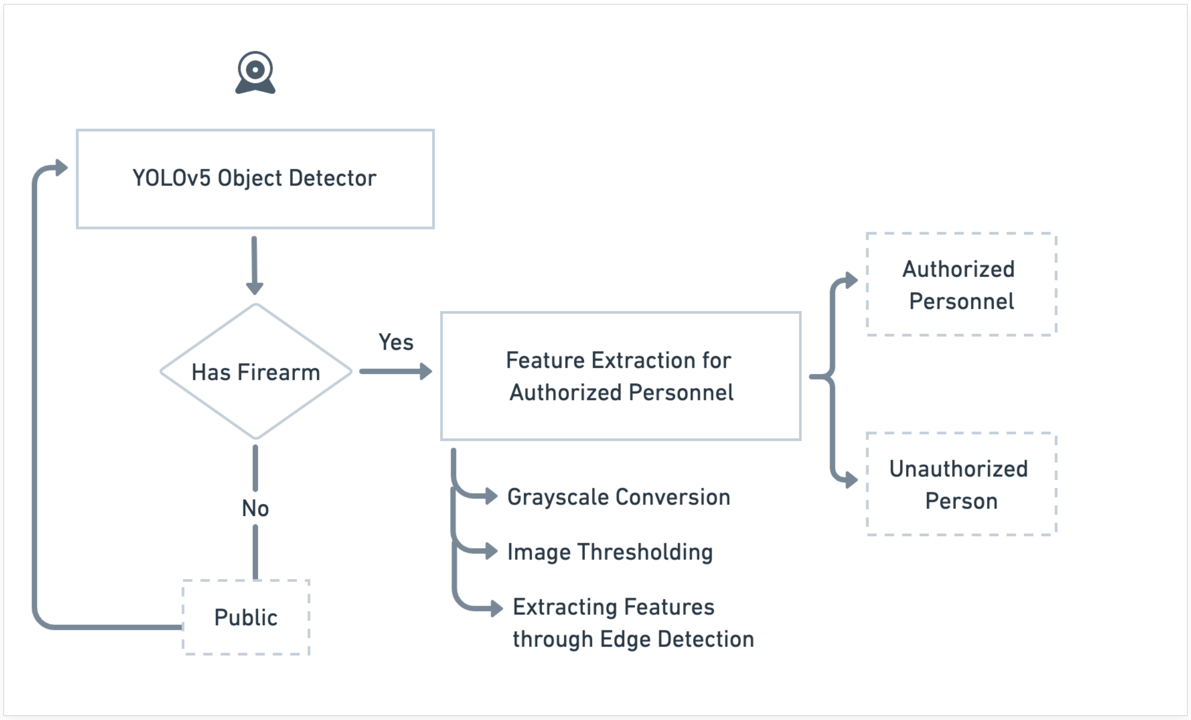

Figure 1: Flowchart

1.2 Input and Outputs

1.2.1 Input

The project will require a collection of image datasets that contain all three classes: Unauthorized Individuals with Firearms, Authorized Individuals with Firearms, and General Public. To identify people with firearms, I have selected the "Youtube-GDD" dataset, which will need to be further annotated to include more information about the person holding the firearm. For the General Public class, I plan to use stock images sourced online. I will carefully choose images where the background environment for the individual with firearms and the general public match closely to ensure that the classifier can correctly identify the relevant features. This serves as the input to my model.

1.2.2 Output

The ultimate objective of this project is to classify individuals detected in video footage as either threats or non-threats, and label them accordingly. In this context, individuals who are authorized to carry firearms would be classified as non-threats, allowing for proper identification of the person of interest.

As an intermediate goal, the project aims to determine whether a firearm is present in the video footage or not.

2 Data

2.1 What data will this project require?

To complete this project, there should be a dataset comprising of three distinct types of images:

1. Images of individuals holding firearms in public areas
2. Images of authorized personnel such as policemen and military personnel holding firearms
3. Images of individuals in public areas such as universities and malls without any firearms.

2.2 Will I be using my own data set or will I use an existing dataset and How do I plan to acquire such data?

I will be using a combination of the YouTube-GDD dataset with my own dataset. This dataset contains approximately 5000 annotated images extracted from high definition YouTube videos tagged as "guns," "shooting," and "weapon show." Subsequently, I will extract the same number of pedestrian images from various online stock image sources to classify them as members of the public.

To further refine the dataset, I plan to use annotating tools like Roboflow and labelImg to classify the annotations as authorized or unauthorized personnel. This will serve as the input for this project

2.3 What is the scale of the data e.g., the number of images needed, etc.

Currently I'm confident that the dataset available with 5000 annotated images is fairly sufficient to build the model. All these images are extracted from high definition youtube videos at a resolution of 640px x 640px. Also, these images will be down-scaled to build the model in a computationally efficient way.

3 Coding Resource Requirements

3.1 What libraries or other code do I plan to use?

For the scope of this project, I plan to use three libraries: opencv-python, PyTorch, and YOLOv5.

opencv-python, is a library commonly used for image and video processing. In this project, I plan to use it to preprocess the images by converting them to grayscale, thresholding them, and producing a binary image suitable for YOLOv5.

PyTorch, is a popular deep learning framework that provides efficient GPU support for training neural networks. I plan to use PyTorch to develop a convolutional neural network (CNN) that will enable image classification and the word prediction model.

YOLOv5, is an object detection algorithm that uses deep learning to detect and classify objects in real-time. I plan to use YOLOv5 in conjunction with PyTorch to create a model for identifying objects in the images.

As the project progresses, additional libraries may be incorporated as needed.

3.2 Whether I intend to utilize any code from an open repository and, if so, what parts of the code you plan to use?

While I don't plan to use complete code from any GitHub repository, I may reference some available code as inspiration for my own implementation.

3.3 What parts/aspects of the project I will code independently?

The majority of the project will be implemented independently coded by myself. Especially the dataset annotation and classifying the individuals as authorized and unauthorized will be coded individually by me.

4 Computational Resource and Effort Requirements

4.1 What computational resources do I plan to use for the project?

I plan to train the model entirely on CPU using my secondary machine. As per my research, YOLOv5 will take around 30 minutes for training at 25 epochs, so I can empirically alter epochs based on my requirement for running 5000 annotated images. Also, If my computational resource doesn't suit the requirement, I might use a subset of the entire data to complete the model.

4.2 What are the estimated man hours I anticipate for the project?

Currently, I don't have a precise estimate for the number of days required to complete the project. However, since I'm the only person working on it, I anticipate investing about an hour per day on average, resulting in an estimated total of 50 man hours.

Please note that this estimate may vary depending on the project's progress and any unforeseen complications that may arise

5 Evaluation

5.1 How will I define the success of the project

The success of the project is defined by the effectiveness in detecting unauthorized individuals with firearms the real-time and tracking them.

5.2 How will I measure this success and provide the evaluation metrics ?

There are several parameters to quantify the accuracy of the model.

1. Average Precision (AP): measuring the accuracy of the model in detecting objects within an image. The metric computes the area under the Precision-Recall curve and can be calculated for each class individually or as an average over all classes.
2. Mean Average Precision (mAP): the average of the AP scores across all the classes. It gives an overall measure of the model's performance.
3. F1-score: the harmonic mean of precision and recall, which are calculated based on the true positives, false positives, and false negatives. It is useful in scenarios where both precision and recall are equally important.
4. Accuracy: measures the percentage of correctly identified objects out of all the objects present in the image.

The success of the project is measured by comparing the achieved the Average Precision(AP) and Mean Average Precision (mAP) values and analysing how close it matches to the state of the art technologies.

6 Project Expectations

6.1 Why am I most excited about for this project?

The most exciting aspects of this project is the potential to make a positive impact on public health by mitigating gun violence. Gun violence is a pressing issue in many countries, and a firearm detection project can be a valuable contribution towards addressing this socially-relevant problem. Additionally, object detection is an essential skill in the field of computer vision and serves as a foundational technique for solving complex challenges. By developing our expertise in object detection, I can further contribute to the advancement of computer vision and its potential applications beyond this specific project.

6.2 What do I hope to learn the most?

By the end of the project, I hope to learn various image processing techniques and how they can be applied sequentially to effectively segment images in classification problems. This knowledge will enable me to make informed decisions when selecting and implementing the most suitable techniques for future image-based projects.

References

[1] Gu Yongxiang, Liao Xingbin, and Qin Xiaolin. Youtube-gdd: A challenging gun detection dataset with rich contextual information. *arXiv preprint arXiv:2203.04129*, 2022.

[2] Mario Alberto Duran-Vega, Miguel González-Mendoza, Leonardo Chang, and Cuauhtemoc Daniel Suarez-Ramirez. Tyolov5: A temporal yolov5 detector based on quasi-recurrent neural networks for real-time handgun detection in video. *CoRR*, abs/2111.08867, 2021.