Universidade de Aveiro
2023

Pedro
Monteiro

Deteção de Armas em Tempo Real em Filmagens de Videovigilância

Real-Time Weapon Detection in Surveillance Video Footages

PROPOSTA
DE TESE

Universidade de Aveiro
2023

Pedro Monteiro

Deteção de Armas em Tempo Real em Filmagens de Videovigilância

Real-Time Weapon Detection in Surveillance Video Footages

PROPOSTA DE TESE

"*In the age of technology, surveillance is not about watching but about predicting human behavior.*"

— Trevor Paglen

Universidade de Aveiro
2023

Pedro Monteiro

Deteção de Armas em Tempo Real em Filmagens de Videovigilância

Real-Time Weapon Detection in Surveillance Video Footages

Proposta de Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à conclusão da unidade curricular Proposta de Tese, condição necessária para obtenção do grau de Mestre em Engenharia Informática , realizada sob a orientação científica do Doutor (Osvaldo Manuel da Rocha Pacheco), Professor Auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro, e do Doutor (Gonçalo Carnaz), Professor auxiliar convidado do Departamento de Electrónica, Telecomunicações e Informática da Universidade de Aveiro.

Texto Apoio financeiro do POCTI no âmbito do III Quadro Comunitário de Apoio.

Texto Apoio financeiro da FCT e do FSE no âmbito do III Quadro Comunitário de Apoio.

Dedico este trabalho à minha esposa e filho pelo incansável apoio.

o júri / the jury

presidente / president — Prof. Doutor João Antunes da Silva
professor associado da Faculdade de Engenharia da Universidade do Porto

vogais / examiners committee — Prof. Doutor João Antunes da Silva
professor associado da Faculdade de Engenharia da Universidade do Porto

Prof. Doutor João Antunes da Silva
professor associado da Faculdade de Engenharia da Universidade do Porto

Prof. Doutor João Antunes da Silva
professor associado da Faculdade de Engenharia da Universidade do Porto

Prof. Doutor João Antunes da Silva
professor associado da Faculdade de Engenharia da Universidade do Porto

Prof. Doutor João Antunes da Silva
professor associado da Faculdade de Engenharia da Universidade do Porto

agradecimentos / acknowledgements

Agradeço toda a ajuda a todos os meus colegas e companheiros.

Palavras Chave

texto livro, arquitetura, história, construção, materiais de construção, saber tradicional.

Resumo

Um resumo é um pequeno apanhado de um trabalho mais longo (como uma tese, dissertação ou trabalho de pesquisa). O resumo relata de forma concisa os objetivos e resultados da sua pesquisa, para que os leitores saibam exatamente o que se aborda no seu documento.
Embora a estrutura possa variar um pouco dependendo da sua área de estudo, o seu resumo deve descrever o propósito do seu trabalho, os métodos que você usou e as conclusões a que chegou.
Uma maneira comum de estruturar um resumo é usar a estrutura IMRaD. Isso significa:

- Introdução
- Métodos
- Resultados
- Discussão

Veja mais pormenores aqui:
`https://www.scribbr.com/dissertation/abstract/`

Keywords textbook, architecture, history, construction, construction materials, traditional knowledge.

Abstract

An abstract is a short summary of a longer work (such as a thesis, dissertation or research paper).
The abstract concisely reports the aims and outcomes of your research, so that readers know exactly what your paper is about.
Although the structure may vary slightly depending on your discipline, your abstract should describe the purpose of your work, the methods you've used, and the conclusions you've drawn.
One common way to structure your abstract is to use the IMRaD structure. This stands for:

- Introduction
- Methods
- Results
- Discussion

Check for more details here:
`https://www.scribbr.com/dissertation/abstract/`

Conteúdo

Lista de Figuras

Lista de Tabelas

Lista de Excertos de Código

Glossário

CAPÍTULO 1

Introduction

A sort description of the chapter.
A memorable quote can also be used.

1.1 Acrónimos

Primeira e seguintes referências: **h2o!** (**h2o!**), **h2o!**
Plural, acrónimo expandido e curto: **h2o!**s, **h2o!**, **h2o!**
Com citação[1]: **adsl!** (**adsl!**), **adsl!**

1.2 Fontes

- Tiny
- Scriptsize
- Footnotes
- Small
- Normal
- large
- Large
- LARGE
- huge
- Huge

1.3 Unidades

Utilizando o pacote `siunitx` é possível utilizar unidades do Sistema Internacional. Exemplo: a aceleração da gravidade é de $9.8\,\mathrm{m\,s^{-2}}$ e um ficheiro ocupa $1\,\mathrm{MiB}$.

[1]Necessária entrada na bibliografia

1.4 Code Blocks

Uma listagem pode ser apresentada com o ambiente `listing`, que é um float (objeto flutuante, tal como uma figura ou uma tabela).

A listagem em Código **??** mostra um exemplo em C.

Código 1: This caption appears below the code.

1.5 Citações

Algumas formas distintas de citar:

- **Apenas referência**: **rfc44**
- **Apenas data**: **rfc44**
- **Apenas ano**: **rfc44**
- **Apenas autor**: **rfc44**
- **Apenas editor**:**rfc44**
- **Autor e referência**:**rfc44**

CAPÍTULO 2

Literature Analysis and State of Art

This chapter delves into the definition and categorization of weapons, while also elucidating the concept and significance of CCTV. Deep learning approaches and algorithms employed for this purpose are thoroughly analyzed. Additionally, the contributions of other authors in this field are dissected, offering a comprehensive review of their methodologies and findings. A significant portion of this chapter is dedicated to examining the datasets used by these scholars, highlighting their relevance, comprehensiveness, and potential limitations in the context of real-time weapon detection.

2.1 Context

2.1.1 Weapons

The Cambridge Dictionary[1] defines a weapon as "any object used in fighting or war, such as a gun, bomb, knife, etc.". This definition suggests that a weapon encompasses any tool or instrument intended to inflict damage or harm, whether on living beings, structures, or systems. Weapons have diverse applications, ranging from hunting and self-defense to warfare. However, their fundamental purpose remains unchanged: they amplify the user's ability to exert force, in either an offensive or defensive capacity.

Weapons can be categorized based on various criteria, including their range, mechanism, or intended use. However, this study will primarily focus on two categories: firearms and melee weapons.

Melee weapons are close-combat instruments that require the user to be in direct proximity to their target, like swords, daggers, maces, and clubs.

Firearms are a subset of ranged weapons that discharge projectiles powered by rapidly expanding high-pressure gas from chemical reactions. They can be further categorized into:

- Handguns: Small, handheld firearms like pistols and revolvers.
- Rifles: Designed for accuracy, rifles have a longer barrel and are often used in situations requiring precision.

[1]https://dictionary.cambridge.org/dictionary/english/weapon

- Shotguns: These fire shells that contain multiple pellets, making them effective at close range.
- Automatic and Semi-Automatic: Automatic firearms continuously fire bullets as long as the trigger is pressed, while semi-automatics require a trigger pull for each shot.

Figura 2.1: Knife Example

Figura 2.2: Different Firearms Example

2.1.2 CCTV

CCTV stands for Closed Circuit Television. It is a video system that consists of strategically placed video cameras around an area that records footage. This footage is then transmitted to display monitors for real-time viewing as well as for playback purposes.

The primary purpose of a CCTV system is to enhance the security of a location. It provides surveillance over key areas continuously. This is especially beneficial for large premises or locations that store valuable equipment, products, or information. Apart from recording video footage, a CCTV system can also send notifications if there's activity or movement detected by a specific camera at a predetermined time, such as during the night when the business premises are closed. Such notifications can be crucial in alerting to potential security breaches. Moreover, while a CCTV system is instrumental in monitoring on-site activity during and outside of working hours, it also aids in identifying wanted criminals and acts as a deterrent to potential intruders.

Figura 2.3: CCTV Cameras Example

2.2 Object Detection Overview

2.2.1 General Concepts in Object Detection

Object detection is a crucial facet of computer vision. At its core, it involves identifying and locating objects within an image. As defined by Roboflow **rfc12**, object detection "identifies objects, and their locations, in an image". A system centered around this concept not only "returns the coordinates of the objects it has been trained to recognize"but also provides a confidence level. This confidence measure indicates the system's assurance in the accuracy of its prediction.

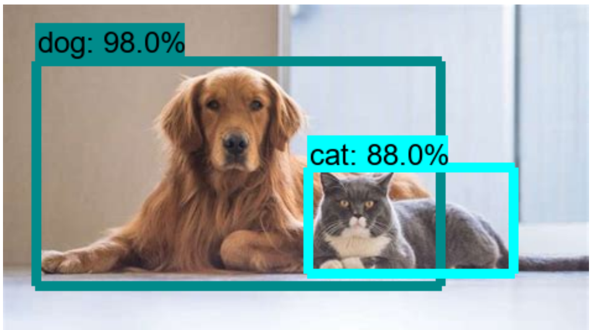

Figura 2.4: Object Detection Image Example **rfc15**

So, object detection involves identifying objects within images. But how are these objects recognized? The key lies in training object detection models. This training process involves feeding a neural network with images of an object presented in various scenarios, such as different backgrounds and angles. Each image is labeled with the corresponding object and its location. As the network processes these images, it learns distinctive features that help in recognizing and locating the object in new, unseen images.

According to V7 Labs **rfc11**, before deep learning took off in 2013, the majority of object detection relied on traditional machine learning methods. These methods identified prevalent features within images and categorized their groupings using tools like logistic regression, color histograms, or random forests. However, contemporary deep learning approaches significantly surpass these in performance.

Object detection can be categorized into two distinct groups (Figure ??):

A. Single-stage object detectors

These detectors process the entirety of an input image in just one pass to predict the presence and location of objects. By bypassing the Region of Interest (RoI) extraction phase, Figure **??**, and directly classifying and adjusting the candidate anchor boxes, they emphasize computational efficiency.

Regarding the concept of Region of Interest, **rfc13** defines it as a "proposed region extracted from the original image", which could potentially contain a relevant object. It's possible to propose thousands of such regions in the RoI extraction phase, depending on the image's content and the extraction algorithm's complexity. However, it's important to note that an RoI is NOT a bounding box. While it might resemble one visually, its function is fundamentally different. An RoI works simply as a suggested area that requires further processing, which is typically done during the RoI pooling stage[2].

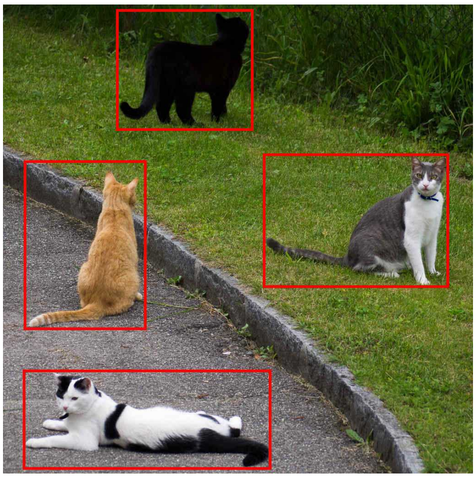

Figure 2.5: Regions of Interest, from Kemal Erdem Article **rfc13**

B. Two-stage object detectors

These detectors undertake a two-step approach in processing images. In the initial phase, extract Regions of Interest (RoIs) to generate a set of potential object locations. The subsequent phase refines these proposals to generate the final predictions.

Single-stage detectors typically prioritize speed and efficiency, making them ideal for real-time applications, whereas two-stage detectors often deliver greater accuracy in complex scenarios, though this method requires more computational power.

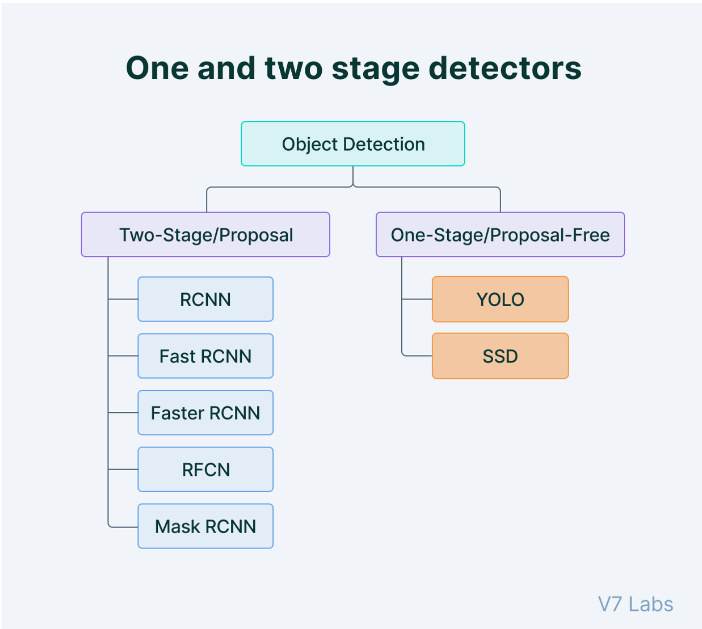

Figure 2.6: One And Two Stage Detectors, from V7 Labs **rfc11**

[2]While RoI extraction is about identifying potential object regions, RoI pooling is about converting these regions into a consistent format suitable for further processing

Due to the variety of existing models and in order to understand which is the most appropriate, there are several metrics that allow determining and comparing the predictive performance of different object detection models. The two predominant metrics are Intersection over Union (IoU) and Average Precision (AP).

Intersection over Union **rfc14**, Figure **??**, is a good metric to calculate localization accuracy and determine localization errors in object detection models. By dividing the intersection by the union, IoU offers the ratio of overlapping area to the entire area, thus effectively indicating the proximity of the predicted bounding box to the true bounding box.

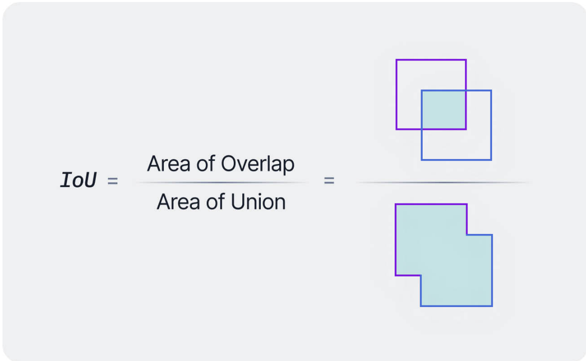

Figure 2.7: Intersection over Union, from V7 Labs **rfc14**

Average Precision (AP) **rfc14** is determined by the area below the precision-recall curve for a set of predictions.

Recall represents the proportion of correctly identified instances out of all actual instances for a given class. On the other hand, precision is the proportion of true positives in relation to all the predictions made by the model.

Recall and precision offer a trade-off that is graphically represented into a curve by varying the classification threshold. The area under this curve produces the Average Precision for each class. Taking the mean of this metric across all classes results in the mean Average Precision (mAP).

2.2.2 Object Detection Techniques Reviews

This section explores the prevailing algorithms and techniques for object detection. To ensure a comprehensive overview, an extensive search was conducted across multiple online repositories and databases, sourcing both pioneering works and recent advancements to provide a balanced and in-depth perspective on the topic. To pursue this objective, the selected reviews[3] are subsequently analyzed in greater detail, delving into their methodologies, findings, and implications for the field, offering a richer understanding of the evolution and current state of object detection techniques.

rfc2 delve into the evolution of detection methods, offering a comprehensive review of deep learning-based object detection frameworks. The review starts with a historical overview of

[3] comprehensive overview or evaluation of existing literature on a particular topic

deep learning, highlighting the role of CNN[4]. The focus then shifts to generic object detection architectures, discussing modifications and strategies to enhance detection performance. Through experimental analyses, various methods for detection tasks are compared, and insightful conclusions are drawn.

rfc8 embark on a journey through the intricate landscape of image processing techniques, presenting a meticulous review of machine learning-powered visual interpretation frameworks. Beginning with the rudiments of computational vision, the study underscores the pivotal role of array-based media computation in today's digital age. The discourse then transitions into a deep dive into prominent image processing algorithms like Single Shot Detection (SSD), Faster Region based Convolutional Neural Networks (Faster R-CNN), and You Only Look Once (YOLO). These algorithms are intricately dissected, revealing the architectural nuances, operational methodologies, and unique features that distinguish them from one another. Through systematic experimentation using datasets like Microsoft COCO, the researchers juxtapose these algorithms, offering profound insights into their individual strengths and caveats.

With their review, **rfc9** underscore the nuances of computer vision, accentuating the transformative role of Deep Convolutional Neural Networks (DCNNs). Their study, while sweeping across various applications from video processing to speech recognition, zeroes in on object detection's pivotal role in fields like transportation and security. They also shed light on pivotal evaluation metrics, notably Average Precision (AP), essential for gauging detector effectiveness. Emphasizing the historical shift from traditional methods to deep learning post-2014, they meticulously dissect frameworks like SSD and YOLO. By juxtaposing these with older techniques and analyzing their performance on datasets like PASCAL VOC and MS COCO, the authors provide invaluable insights into the ever-evolving landscape of object detection.

rfc10 delve deep into the expansive realm of deep learning (DL), offering a comprehensive review of its concepts, convolutional neural network (CNN) architectures, challenges, applications, and prospective directions. The paper starts by highlighting the rising dominance of the DL paradigm, asserting its position as the "Gold Standard" in the machine learning community. Notably, the authors elucidate the unparalleled capability of DL to learn from massive datasets, outpacing traditional approaches in various applications. Venturing further, the discussion unfolds around pivotal CNN architectures, shedding light on their design intricacies and operational mechanisms. The challenges encountered in DL, be they computational or conceptual, are thoroughly explored. Additionally, the authors venture into the myriad of applications where DL has showcased exceptional performance, matching or even surpassing human capabilities.

2.3 Related Work

In the pursuit of a deeper understanding and improvement in weapon detection, numerous pivotal studies have emerged, highlighting various techniques, methodologies, and datasets.

[4]Convolutional Neural Networks

This section presents a thorough overview of these significant contributions, organizing the discussion into three core subsections: research methods, other author's articles, and datasets.

The initial part of the exploration dives into the research methods. This subsection details the approach taken to locate and identify pertinent literature. Given the vast amount of available works, a strategic filtration process was applied to select articles that are of paramount relevance to the topic of weapon detection.

Moving to the core of this section, a detailed examination of the articles[5] is undertaken. Each article is analyzed with respect to its primary objectives, methodologies, and outcomes. In an endeavor to furnish readers with a granular understanding, the techniques and architectures proposed by various scholars are further dissected.

Lastly, the focus shifts to the datasets[6]. This part furnishes a detailed overview of the datasets employed in weapon detection research. The narrative explores the origins of these datasets, the methods used for their sourcing, and underscores their distinguishing features.

2.3.1 Research Method

To initiate this study, a comprehensive review of existing research and projects in this field, with analogous objectives, was imperative. To facilitate this literature survey, the Scopus[7] database, renowned for its vast collection of academic and research publications, was employed.

Due to the vast number of articles across diverse disciplines, it was necessary to make a careful selection. Consequently, specific keywords were utilized to pinpoint articles pertinent to the dissertation's subject matter.

So, utilizing Scopus, a search was executed with the query: ("deep" AND "learning" AND "real" AND "time" AND "weapon" AND "detection"). From the 70 records retrieved, a meticulous selection process resulted in 8 articles that align closely with the theme of this study, as illustrated in Figure ??.

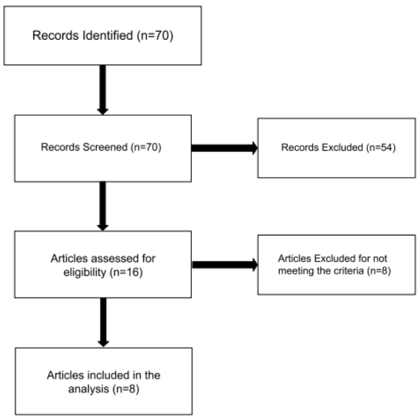

Figure 2.8: Scopus Query, Author's own

[5]detailed written discourse on a specific subject
[6]collection of related data points, often organized in tables, arrays, or other structures.
[7]`https://www.scopus.com/home.uri`

2.3.2 Articles

rfc3 address the pressing concern of rising incidents involving firearms and knives due to insufficient security checks. Noting that while CCTVs have become prevalent, their constant surveillance demands often surpass human monitoring capabilities. The paper primarily introduces an automated weapon detection system, which employs the YOLOv5 deep learning model, tailored to a specially curated dataset. This innovation excels in the detection of various firearms and knives, showcasing an impressive F1 score of 0.95 when applied to CCTV footage. The study underscores the potential of neural networks and AI in enhancing real-time security measures.

In a world increasingly focused on security and safety, **rfc4** emphasize the vital role of establishing a secure environment to bolster economic growth, especially for investors and tourists. While CCTV cameras have been a cornerstone for surveillance, their dependency on human monitoring highlights a gap in ensuring continuous vigilance. Recognizing the challenges of real-time weapon detection — such as varied viewing angles, occlusions, and carrier concealment — the research undertakes the task of leveraging advanced, open-source deep learning algorithms to detect potential threats. Given the absence of a standardized dataset, the researchers meticulously curated their own, drawing from diverse sources, including original photos, online repositories, and even film databases. Their exploration spanned several algorithms, like VGG16, Inception variants, and YOLO iterations. Through rigorous testing prioritizing precision and recall over mere accuracy, YOLOv4 emerged as the frontrunner, achieving a notable F1-score of 91% and a mean average precision of 91.73%, setting a new benchmark in the field.

In the research conducted by **rfc5**, the growing menace of illicit firearms in public spaces despite the omnipresence of CCTVs is highlighted. While CCTVs offer visual insights, the human element in consistently monitoring vast streams of footage often leads to oversights, especially in detecting concealed weapons. Addressing this gap, the study introduces an advanced weapon detection methodology leveraging the YOLOv5 deep learning framework. This model is fine-tuned to a dataset comprising 1496 images and demonstrates a superior ability to detect handguns in varied scenarios. Notably, the system achieved an impressive mean average precision (mAP) of 91.76% when tested. A comparative analysis, Figure **??**, with the YOLOv3 model revealed that YOLOv5 not only has enhanced accuracy in handgun detection but also offers faster detection speeds and is computationally less demanding.

Object Detection	**mAP (%)**	**Detection Speed per image (s)**	**No. frames per second**	**No of Parameters (million)**	**Model Size (MB)**
YOLOv3	78.90	0.054	13	7.5	236
YOLOv5	91.76	0.023	24	62	14

Figure 2.9: rfc5 Models Performance Comparison

In the context of Smart Cities, **rfc18** address the urgent need for enhanced urban security through advanced surveillance. While CCTVs are prevalent, the sheer volume of visual data often challenges human analysis. The authors present a novel approach, combining super-resolution techniques with the YOLOv4 deep neural network, to detect knives in complex images, a task made difficult by variables like shape, size, and lighting. Their research showcases the potential of AI surveillance to enhance real-time security, making video footage not just viewable but actionable and quantifiable. This study sets the stage for future extensions, such as firearm detection, and works as a testament to the transformative power of AI in urban safety protocols.

In the study by **rfc17**, the escalating urban crime rates, particularly involving firearms and sharp objects, are brought to the forefront. While CCTVs are omnipresent in modern cities, the vast amount of data they produce often overwhelms manual surveillance efforts. Addressing this gap, the authors have developed a cutting-edge solution using the YOLOv8 deep learning framework, fine-tuned on a custom dataset comprising images of pistols, knives, and screwdrivers. The system, Figure **??**, not only identifies weapons but also evaluates potential threat levels by analyzing the arm position of the person in possession. Demonstrating a notable 93% accuracy on CCTV recordings, the study exemplifies the transformative role of AI in enhancing urban security protocols.

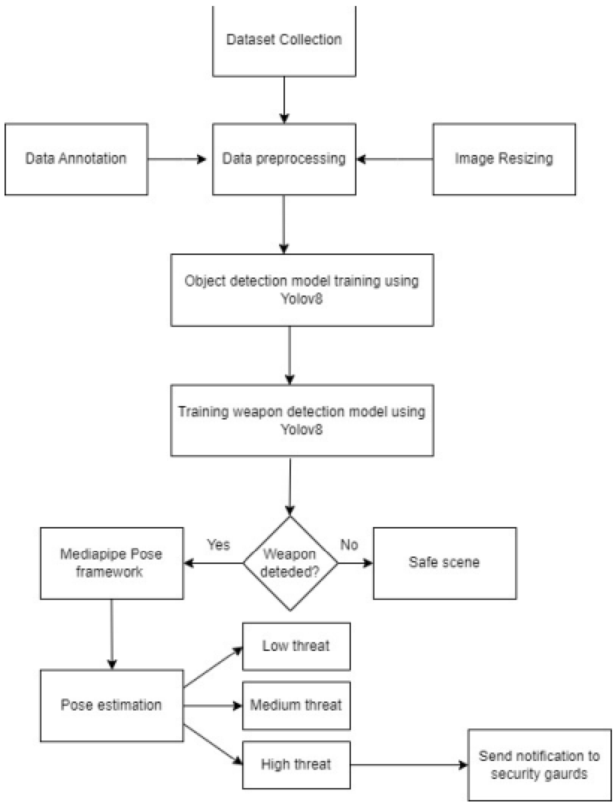

Figure 2.10: rfc17 Solution Architecture

rfc6 research tackles the increasingly alarming scenario of weapon-related threats in crowded public spaces like banks, airports, and railway stations. Recognizing that while the deployment of CCTV cameras has seen a global surge, the sheer volume of footage can overwhelm human operators. This paper introduces a sophisticated weapon detection system, Figure **??**, that uses the power of deep learning, specifically through the use of Convolutional Neural Networks (CNN). Trained on a comprehensive dataset of 10014 images, the model adeptly identifies knives, small guns, and long guns, achieving an admirable accuracy rate of approximately 85%.

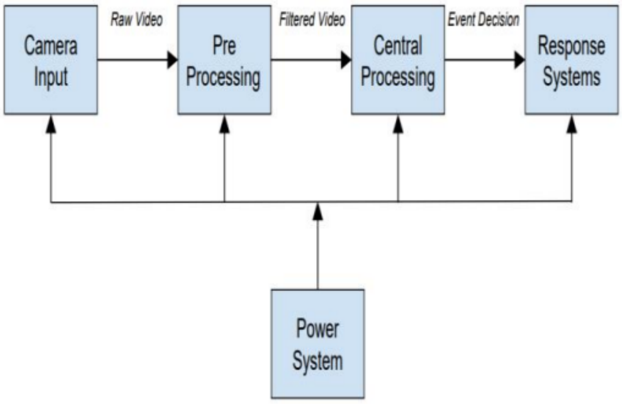

Figure 2.11: rfc6 Architecure Diagram

rfc19 introduce a forward-thinking security system architecture that uses deep learning and image-processing for real-time weapon detection in CCTV feeds. By periodically capturing images from these feeds and analyzing them using a convolutional neural network (CNN), the system can quickly identify potential threats. Upon detection, security personnel are promptly alerted via a mobile application, accompanied by an image of the potential threat. Impressively, the system claims a 92.5% accuracy rate and can detect threats in a mere 1.6 seconds, underscoring the transformative potential of integrating AI with security protocols.

rfc20 highlight that while CCTV systems are everywhere, their efficacy is often hampered by the reliance on police personnel for monitoring. Addressing this challenge, the study introduces an integrated weapon detection system for CCTVs. Drawing from two public datasets, ARMAS and IMFDB, the research employs various object detection techniques, including SSD MobileNet-V1, EfficientDet-D0, and notably, Faster R-CNN Inception Resnet-V2. Specifically, the Faster R-CNN Inception V2, when used with the ARMAS dataset, achieved a mAP of 0.540, and Average Precision scores of 0.793 at 0.5 IoU and 0.627 at 0.75 IoU.

rfc7 emphasize the alarming rise in criminal activities in urban spaces, despite the ubiquity of CCTV systems in public areas. Recognizing the limitations of human supervision for these vast surveillance networks, they introduce an intelligent crime detection mechanism, Figure **??**. Utilizing the SSD Mobilenet architecture fine-tuned on the DaSCI Weapon Dataset, their solution adeptly identifies a broad spectrum of weapons, yielding an accuracy rate of 81%. Moreover, with the integration of the GRU-based behavior analysis model, the system achieved a notable accuracy of 95.97% in detecting suspicious actions.

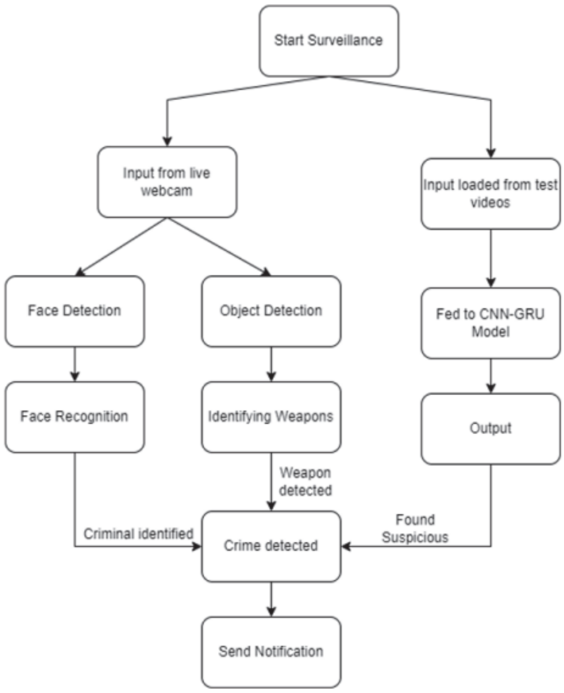

Figure 2.12: rfc7 System Architecture

2.3.3 Datasets

In Rehman's study **rfc3**, the dataset comprises three distinct classes: Short Guns, Long Guns, and Knives - Figure **??**. This dataset is gathered from a variety of platforms, including surveillance videos from YouTube, simulations of firearms and knives created within the Unity framework, and images of pistols, revolvers, and knives from the Open Images Dataset as well as Kaggle. For video sources, individual frames were extracted and subsequently labeled manually using the LabelImg Python utility. To enhance the model's training performance, data augmentation techniques were employed.

Short guns	Long guns	Knives	Total
4370	3426	1984	10567

Figure 2.13: rfc3 Dataset Images Classes Number

rfc4 focused on creating robust datasets for real-time weapon detection, specifically targeting the recognition of pistols within various environments. The data was meticulously gathered from diverse sources including the internet, YouTube CCTV videos, GitHub repositories, research groups, and the Internet Movie Firearm Database (imfdb.org). They constructed three distinct datasets, being categorized into 'Pistol' and 'Not-Pistol' classes. The 'Pistol' class included pistols, revolvers, and other similar handheld weapons, while the 'Not-Pistol' class comprised objects often mistaken for weapons, such as wallets, cell phones, and metal detectors. This deliberate inclusion of 'confusion objects' in the 'Not-Pistol' class was a strategic approach to reduce false positives and negatives, thereby enhancing the precision and accuracy. For the data preparation and model training the researchers implemented data pre-processing steps to standardize the dataset, ensuring consistency in image size and resolution, and applied mean normalization. Also, used data augmentation strategies to artificially expand the dataset, improving the model's ability to generalize.

In the study conducted by **rfc5**, models were trained using Google Collab, both Yolov3 and Yolov5, utilizing pre-trained weights on the COCO dataset **rfc16**. The dataset, comprising 1496 images, was strategically segmented into training, validation, and testing sections. Out of these, 1200 images were allocated for training, while the remaining 296 were divided for both validation and testing, each one with 148. Training a model for 40 epochs took around 120 minutes, with a standard image size set at 416x416 pixels. Both YOLOv3 and YOLOv5 were trained within the PyTorch framework. The backbone feature extraction network, used across all experiments, was pre-trained using the COCO dataset. Additionally, fine-tuning of the backbone was allowed during training to enhance the representation of handguns.

rfc6 used a comprehensive dataset consisting of 10,014 distinct weapon images to train and evaluate the model's accuracy. These images were categorized into three primary object types: knives, with a total of 3,641 images; long guns, comprising 2,497 images; and small guns, which included 3,876 images. The dataset was strategically divided, allocating 80% of the images (8,011 images) for training and the remaining 20% (2,003 images) for testing. Each image used had a resolution of 240 x 240 pixels, and during the training phase, they were processed in batches of 32.

rfc7 employed two distinct datasets to train and evaluate the models. The first, the DaSCI Weapon Dataset, stands out in the realm of weapon detection for its ability to distinguish between weapons and commonplace objects. This dataset facilitated the training of the system to detect approximately 91 weapons of mass destruction. For the Suspicious Behavior Prediction module, the CAVIAR Dataset was utilized. This dataset comprises video clips representing six varied human actions. Annotations for these .mpg videos are stored in XML format, detailing both the role and context associated with each action. This information is crucial in determining whether a particular action can be classified as suspicious. To process this data, the XML file was parsed, aligning annotations with individual video frames. Each video was deconstructed frame by frame, extracting relevant data from the XML annotations. Subsequently, the data was labeled either as suspicious or not suspicious.

rfc18 used two primary datasets to training and testing models related to knife detection in surveillance systems. The DaSCI Dataset focuses on knife detection, comprising 2,078 images with various visual features of knives. The images were mainly sourced from the internet, including YouTube videos. On the other hand, the MS COCO 2017 Dataset is a broader computer vision dataset with 330,000 images across 80 classes. Within this dataset, 4,326 images are labeled for the 'knife' class.

rfc17 faced challenges in gathering a comprehensive weapon dataset, particularly for knife and screwdriver images, due to the lack of labeled datasets suitable for classifier training. To address this, they employed web scraping techniques, sourcing images from various websites and GitHub repositories. Following the collection process, manual labeling of the images was undertaken using graphical annotation tools such as LabelImg and Roboflow. The images were then preprocessed to achieve a standardized resolution of 416 x 416. The final dataset comprised approximately 6,000 images: 2,000 each of pistols, knives, and screwdrivers. For model training and testing, the images were split in an 80:20 ratio, respectively.

rfc19 set up a Raspberry Pi B+ and camera 1.8m high, facing a lab entrance 4.5m away to capture images of consenting students entering and exiting. To simulate potential threats, students were provided with handgun replicas, with minimal guidance on how to hold them, resulting in a variety of poses. The dataset started with 14k images at 1920p x 1080 resolution. After data augmentation like scaling and rotation, it expanded to 28k images. A challenge was the gun's small size in images, increasing misclassification risks.

rfc20 sourced images of firearms-related crime intentions from two primary datasets: the ARMAS Weapon Detection Dataset, which contains 3,000 clear images of pistols, some held by individuals; and the IMFDB Weapon Detection System, which combines 4,940 diverse images of pistols and rifles from the Internet Movie Firearms Database and some CCTV footage. After collecting the data, the images were preprocessed to fit object detection training protocols and subsequently divided into training and testing subsets using an 80:20 ratio.

2.4 Research Results

2.4.1 Weapon Detection Deep Learning Algorithms

2.4.2 Weapon Detection Datasets