Concealed Object Detection on Terahertz Videos using YOLOv5m

PRE-DISSERTATION REPORT SUBMITTED
FOR THE AWARD OF DEGREE OF

M. Sc. (DATA SCIENCE)

BY

LAKSHMY SANTHOSH
Roll No: 222828

UNDER THE GUIDANCE OF

PROF. SINGARA SINGH

COMPUTER SCIENCE AND INFORMATION TECHNOLOGY
CENTRAL UNIVERSITY OF HARYANA
MAHENDERGARH, HARYANA, INDIA-123031
DECEMBER 2023

Certificate

I hereby declare that the work being presented in this thesis, in fulfillment of the requirements for the award of degree of **M. Sc. (Data Science)** submitted in Department of Computer Science and Information Technology, Central University of Haryana, is an authentic record of my own work carried out under the supervision of Dr. Singara Singh, Professor, Department of Computer Science and Information Technology and refers other researcher works which are duly listed in the reference section. The matter presented in this thesis has not been submitted for the award of any other degree of this or any other university.

(**Lakshmy Santhosh**)
Registration No. 222828

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge and belief.

(**Dr. Singara Singh**)
Professor
Computer Science and Information Technology
Central University of Haryana, Mahendergarh
Supervisor

Acknowledgement

I extend heartfelt gratitude to my supervisor, **Dr. Singara Singh**, for their unwavering support, steadfast guidance, and invaluable mentorship during the pre-dissertation phase. Their profound expertise and continuous encouragement have significantly shaped my research endeavors. I would also like to convey my appreciation to the Head of the Department (HOD), **Dr. Keshav Singh Rawat**, for providing me with the transformative opportunity to embark on this academic journey. Their unwavering belief in my capabilities has been a powerful and motivating force throughout this process. Additionally, I am grateful for the myriad of online resources that have expanded my knowledge, facilitating seamless access to a wealth of information. The collective support from these individuals and resources has played a pivotal role in the profound development of my pre-dissertation work, and I am sincerely thankful for their contributions to my academic journey.

(**Lakshmy Santhosh**)

Abstract

Exploring the domain of concealed object detection, this preliminary dissertation work utilizes the YOLOv5m model, celebrated for its enhanced flexibility in real-time applications. The model's architecture, encompassing the stem, CSPDarkNet53 backbone, neck, and head, is meticulously crafted for processing tensor-formed images and initiating the object detection process. The stem module engages in initial preprocessing and feature extraction, while the CSPDarkNet53 backbone extracts pivotal high-level features for object identification. Represented by the Path Aggregation Network (PANet), the neck refines feature integration and spatial dimensions. Multiple detection heads then predict bounding boxes, objectness scores, and class probabilities at various scales. The outlined methodology employs a publicly available terahertz video dataset for concealed object detection, utilizing the YOLOv5m model. The research strives for heightened accuracy in detecting both dangerous and non-dangerous objects within the terahertz spectrum, encompassing knives, bombs, guns, cigarette boxes, and A4 paper. The dataset's diversity ensures a robust evaluation, with preliminary results indicating promising accuracy in concealed object detection. This foundational work sets the stage for advancing security surveillance through innovative feature extraction techniques, paving the way for future developments.

Abbreviations

YOLO	You Only Look Once
CSPnet	Cross Stage Partial Network
PANet	Path Aggregation Network
AMMW	Active Millimeter Wave
RGB	red, green and blue
M16	Military rifle model 16
CNN	Convolutional Neural Network
AK	Avtomat Kalashnikova(gun)
TT	Tula Tokarev (pistol)
F1	F-1 granade
RGD5	Ruchnaya Granata Distantsionnaya (Hand Granade Remote)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The field of object detection plays a pivotal role in enhancing security surveillance systems, particularly when it comes to identifying concealed and potentially dangerous objects such as guns, explosives, and knives. While the detection of visible objects is relatively straightforward, the challenge becomes more complex when these items are hidden within clothing or bags. To address this challenge, a variety of imaging technologies are employed, including active millimeter wave (AMMW), terahertz (THz), RGB, and thermal imaging.

Active millimeter wave (AMMW) methods leverage deep learning models for explosive detection, but they are often constrained by low spatial resolution and face difficulty in detecting small objects. On the other hand, terahertz (THz) security inspection methods excel in detecting both non-metallic and metallic objects. The shorter wavelength of terahertz waves provides high-precision video, making them effective for identifying concealed objects beneath clothing. Despite its potential, terahertz video surveillance is underutilized in real-world applications.

Thermal infrared imaging, which relies on thermal contrast, offers high spatial resolution and effectively mitigates thermal effects. This technology enables the identification of suspicious items carried by individuals. The integration of terahertz technology has not only elevated video quality but has also accelerated the overall detection process. Various techniques have been developed for object detection, each with its unique advantages. The 2D-discrete wavelet transform (DWT) model, for instance, demonstrates exceptional accuracy. Additionally, deep learning algorithms, especially

those utilizing convolutional neural networks (CNNs), have become prominent in the field of object detection. Optimization strategies, including genetic algorithms and swarm intelligence techniques, contribute to swift and accurate object detection, further enhancing the capabilities of surveillance systems.

In the context of bolstering security surveillance systems and preventing suspicious activities, the study introduces an innovative approach by putting forth a YOLOv5m model for concealed object detection. The principal focus of the model is to elevate the effectiveness of identifying concealed objects with an emphasis on both efficiency and accuracy. This model is structured around three pivotal sections, each meticulously designed to contribute to the overall enhancement of concealed object detection.The initial section, referred to as the stem module, plays a crucial role in the preprocessing and feature extraction processes. This module serves as the foundation for the subsequent stages, where it carefully processes input data and extracts essential features. The efficiency of concealed object detection is significantly amplified through this meticulous preprocessing and feature extraction, setting the stage for robust subsequent analysis.

Following the stem module, the backbone module comes into play, dedicated to extracting high-level features. This section of the model is designed to discern intricate patterns and complex information from the preprocessed data. By delving into the nuances of the input, the backbone module aims to capture critical features that are indicative of concealed objects. This hierarchical feature extraction process is instrumental in elevating the model's ability to discern concealed objects within a given surveillance scenario.

The third and integral component of the model is the neck section, tasked with predicting object locations. Leveraging the features extracted by the backbone module, the neck section engages in a sophisticated analysis to pinpoint the precise locations of concealed objects. Through a meticulous localization process, this section contributes significantly to the overall accuracy of the detection model. The collaborative efforts of the stem, backbone, and neck sections culminate in a comprehensive architecture geared towards optimizing the detection of concealed objects in security surveillance systems.

This multi-section architecture is not merely a concatenation of disparate components; rather, it

is a carefully orchestrated symphony of modules, each playing a distinct yet harmonious role. The synergistic integration of these sections is a testament to the commitment to advancing concealed object detection capabilities in security applications. By intricately weaving together preprocessing, feature extraction, and object localization, the YOLOv5m model stands poised to set new benchmarks in the realm of surveillance technology, fortifying security measures and contributing to the prevention of suspicious activities.

1.1 Importance of YOLOv5

YOLOv5, or "You Only Look Once version 5," has emerged as a pivotal advancement in the field of computer vision and object detection. The significance of YOLOv5 lies in its ability to provide real-time and highly accurate object detection in diverse scenarios. The model's architecture is designed to process images in a single pass, making it exceptionally fast and efficient. This real-time capability is crucial in applications such as autonomous vehicles, surveillance systems, and augmented reality, where rapid and accurate detection of objects is paramount for decision-making and user experience. YOLOv5's superior performance in terms of speed and accuracy has made it a preferred choice for many computer vision researchers and developers, contributing to the advancement of various technologies that rely on robust object detection.

Another key aspect of the importance of YOLOv5 is its open-source nature and its active community support. YOLOv5 is available as an open-source project, allowing researchers and developers worldwide to access, modify, and contribute to its development. This collaborative effort has led to continuous improvements, bug fixes, and the incorporation of state-of-the-art techniques, keeping YOLOv5 at the forefront of object detection research. The open-source nature fosters innovation and democratizes access to advanced computer vision capabilities, enabling a broader range of applications and encouraging the development of novel solutions across industries. YOLOv5's significance extends beyond its technical capabilities, as it serves as a catalyst for progress and collaboration in the rapidly evolving field of computer vision.

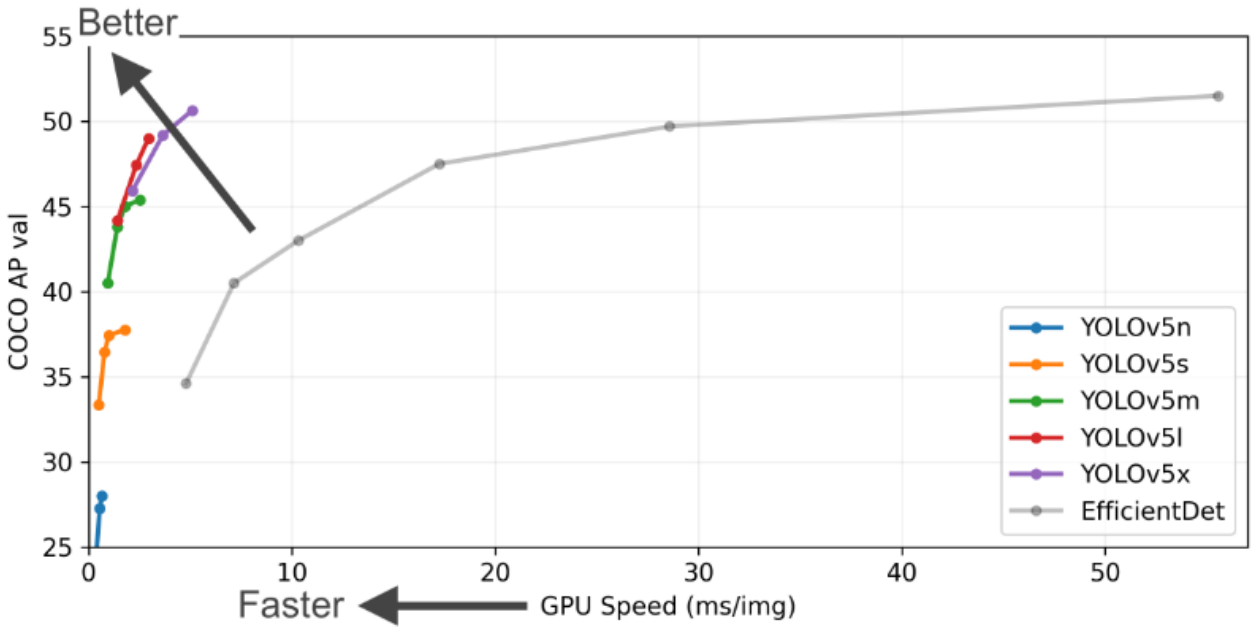

Figure 1.1: Comparison of YOLO Models

1. **COCO AP val (mAP@0.5:0.95):**

   The COCO AP val metric signifies the mean Average Precision across a range of IoU thresholds (0.5 to 0.95). The evaluation is conducted on the extensive COCO val2017 dataset, comprising 5000 images. What sets this analysis apart is the exploration of various inference sizes, spanning from 256 to 1536. This nuanced approach aims to uncover the YOLOv5m model's proficiency in object detection across diverse scales, ensuring a robust assessment of its accuracy.

2. **GPU Speed (AWS p3.2xlarge V100 Instance, Batch-size 32):**

   The GPU Speed metric delves into the model's real-world efficiency by measuring the average inference time per image on the COCO val2017 dataset. Leveraging the computational power of an AWS p3.2xlarge V100 instance, a high-performance GPU configuration, the YOLOv5m model's responsiveness is assessed under the demanding condition of a batch-size of 32. This metric provides valuable insights into the model's practical applicability, particularly in scenarios where quick and efficient object detection is paramount.

1.2 Terahertz Videos

Terahertz (THz) technology has gained attention in the field of security due to its ability to penetrate certain materials and detect hidden objects or substances. Terahertz radiation falls between microwave and infrared in the electromagnetic spectrum, and it has unique properties that make it suitable for security applications.

Terahertz videos for security purposes may involve imaging techniques that utilize terahertz waves to reveal concealed objects or substances beneath clothing or within various materials. These videos can be part of security systems deployed in airports, public transportation, or other secure environments. The terahertz technology can detect items such as weapons, explosives, or drugs that may not be easily identified using traditional screening methods.

Terahertz technology operates in both active and passive modes, with the passive mode being safe for humans and playing a critical role in concealed object detection. The passive terahertz mode is especially advantageous in situations where minimizing human exposure to radiation is essential. This characteristic makes it suitable for applications like security screening, ensuring concealed object detection without posing harm to individuals. The technology has proven its ability to improve video quality, aiding in the rapid and efficient identification of potentially threatening items in various security-sensitive environments. Despite these advancements, the widespread adoption of terahertz video surveillance in diverse real-world settings remains a challenge that requires further exploration and integration.

Chapter 2

Review of Literature

Ensuring public safety requires the effective detection of concealed dangerous objects, including weapons and explosives, especially in crowded locations like airports, train stations, and public events. However, the detection of hidden objects poses a significant challenge due to their small size and diverse shapes. Traditional approaches, such as conventional image processing and computer vision techniques, have limitations in terms of both accuracy and speed. Hence, there is a crucial need to devise a method that enhances the accuracy and robustness of detection for improved public safety.

2.1 Concealed Object Detection

In their work , Kowalski [3] introduced the YOLO3 algorithm designed for recognizing and detecting concealed objects in infrared and terahertz images. The image undergoes processing through a residual block and subsequently through Darknet-53, where features are extracted. Logistic regression is employed for calculating each bounding box, and k-clustering is utilized to determine priors within groups, applying feature mapping to each group. The computational complexity is reduced through the use of binary entropy loss. The classification of regions within the image is accomplished using ResNet.

The paper [5] introduces a normalized accumulation map-based training mechanism for concealed object detection in millimeter-wave images. By calculating the average of binary masks representing object locations, the proposed mechanism assigns different weights to frequently-appeared concealed

objects during confidence loss computation. Experimental results on a millimeter-wave security image dataset demonstrate a 4.43% improvement in mean average precision when applying this approach to the YOLO-v2 object detection network. This innovative training strategy holds promise for enhancing the accuracy of concealed object detection systems in millimeter-wave imaging for security applications.

A Self-Paced Feature Attention Fusion Network (SPFAFN) is proposed [6]. It integrates features of different scales, employs a hierarchical pyramid attention mechanism, and utilizes self-paced learning. Validated on AMMW and passive millimeter-wave datasets, the method outperforms state-of-the-art approaches, demonstrating superior Average Precision (AP) and enhancing concealed object detection in security applications.

Terahertz security screening cameras (TSSCs) offer low radiation and efficient detection but often provide only rough object locations. This paper [7] introduces a convolutional neural network (CNN) approach for automatic object detection and recognition in THz security image sequences. The method employs sparse and low-rank decomposition (SLD) for rough detection, refining locations through shape knowledge and morphological processing. Detailed recognition utilizes supervised training with the Faster R-CNN model, enhancing efficiency by a narrow-band approach. Extensive experiments validate high-performance results in terms of accuracy and efficiency, promising advancements in terahertz security screening applications.

The research conducted in this work [2] leveraged a novel and innovative approach by integrating the Mutation-Enabled SWALP (Swarm Algorithm with Local Parallelism) algorithm in conjunction with the widely acclaimed YOLOv5 (You Only Look Once) object detection framework. This fusion of cutting-edge technologies aimed to address and enhance various aspects of computer vision, particularly in the domain of real-time object detection and tracking.

2.2 Limitations

Some of the limitations of the works in the field of concealed object detection using algorithms like YOLO, CN, etc. are listed:

- Precision Challenges in Object Localization: Achieving precise object localization, especially amidst noise and clutter, is a critical challenge for detection algorithms.
- Need for Abundant Training Data: Deep learning methods often demand extensive annotated training data, which is limited, especially for specific concealed object types.
- Accuracy Struggles Due to Object Variation: The diverse sizes, colors, and shapes of concealed objects create difficulties in their accurate detection.
- Complexity and Computational Demands: Some deep learning approaches, like YOLO3 and Darknet, have high computational complexity, limiting real-time detection in resource-constrained environments.
- Constraints in Image Quality Enhancement: Improving the quality of terahertz or infrared images used in concealed object detection is challenging, with existing methods having limitations.
- Challenges with Small Object Detection: Deep learning struggles to accurately detect small-concealed objects, affecting overall detection performance.
- Accuracy and Robustness Issues in Traditional Approaches: Traditional image processing and computer vision methods lack the accuracy and robustness of deep learning, particularly in complex scenarios and varied object appearances.

2.3 Challenges and Research Gaps

- Need for Speedy Recognition Algorithms: Existing terahertz video detection methods lack fast recognition algorithms crucial for timely threat identification and prevention, posing a gap in real-time applications and response efficiency.
- Research Gap in Terahertz Video Detection: Despite the emergence of terahertz video in security, there's a lack of research on tailored methods for concealed object detection, emphasizing the need for more exploration in this specific area.

- YOLO Parameter Optimization Gap: The use of YOLOv5m in terahertz video detection lacks research on optimizing YOLO parameters specifically for terahertz-based scenarios, requiring further investigation to achieve optimal performance.

- Shortage of Terahertz Video Datasets: The absence of annotated datasets designed for concealed object detection in terahertz video hampers the development and evaluation of effective detection algorithms, highlighting a need for comprehensive and diverse datasets.

- Incomplete Method Comparison: While the proposed method claims superior accuracy over existing ones like CNN, YOLO3, etc., there's a research gap in a comprehensive comparison and evaluation specifically for terahertz-based concealed object detection, necessitating further validation and research.

Chapter 3

Research Methodology

1. The first step was to find a good terahertz dataset. Found a publicly available terahertz dataset from fullvision.ru[4]. Since terahertz datasets are very scarcely available publicly, it was a major part to find a dataset.

2. The utilized dataset originates from a publicly accessible terahertz video dataset generated through a dedicated software developed in the Actor Prolog programming language. The videos are encoded in AP2j format, necessitating the extraction of frames from all 32 videos, each comprising approximately 350 frames.

3. Subsequently, the YOLOv5m model was selected for further analysis.

4. YOLOv5m requires a specific dataset format, including a .txt file containing annotations of bounding boxes for objects within the images. To achieve this, images were annotated using the roboflow website [1]. Due to the time-intensive nature of this process, annotations were created for 40 images from each video.

5. Additional data augmentations, such as flipping and rotation, were introduced to enhance the diversity of the dataset. Following this, the dataset was partitioned into training, testing, and validation sets.

6. The YOLOv5m model was then trained on the prepared dataset, yielding noteworthy results.

Chapter 4

Data Analysis and Interpretation

4.1 Dataset Aquisition

The research endeavors begin with the acquisition of a publicly available terahertz video dataset [4]. This dataset comprises a diverse range of objects, encompassing potential threats such as knives, bombs, and guns, alongside non-dangerous items like cigarette boxes and A4 paper. The detailed class distribution is outlined in Table 3.1. This dataset's diversity lays a robust foundation for a comprehensive evaluation of the concealed object detection system's performance. The inclusion of various object classes ensures the versatility required to assess the model's efficacy across a spectrum of concealed items.

This dataset, obtained from real-world scenarios, provides a rich source for training and evaluating the concealed object detection system, ensuring its effectiveness in diverse security applications.

4.1.1 Dataset Details

The dataset employed in the research is sourced from a comprehensive Terahertz video dataset [4], consisting of 32 distinct videos. These videos are encoded in the AP2J format, necessitating the use of specialized software for viewing and manipulating data properties. The dataset's intricacies involve the actor prolog programming system, developed within the AP2J model specifically for recording video files. Each video record incorporates diverse modalities, featuring several video

Table 4.1: Object Classes in the Terahertz Video Dataset

Class	Object Type	Description
1	A4paper	paper
2	AK	gun
3	AK_noMagazine	gun
4	M16	gun
5	Tin	container
6	axe	-
7	beltholster	pistol
8	bottle	container
9	candyboxLid	-
10	cigaretteBox	-
11	fomka	-
12	glassjar	container
13	hammerAndSickle	-
14	handGranade	granade
15	knife	-
16	meatKnife	-
17	phoneNokia	phone
18	phoneXiaomi	phone
19	pistol	pistol
20	saucepanLid	-
21	shoulderholster	pistol
22	usbDisk	-

streams. Notably, the organization of video files involves grouping four files together, forming a cohesive unit within the dataset structure. This meticulous dataset curation provides a rich and diverse set of scenarios, laying a robust groundwork for exploration into terahertz-based concealed object detection.

4.1.2 Dataset Preparation

The dataset construction involved the utilization of Multimedia_04_3D_normalizer, used in extracting terahertz image frames from the diverse set of 32 videos. An interesting facet of the approach was consolidating videos depicting the same object in various anatomical locations into a unified class, ensuring a coherent representation. For the preprocessing and annotation phase, we leveraged Roboflow [1], a robust platform that streamlined these tasks efficiently.

The resulting dataset encompasses 24 distinct classes, capturing the diversity inherent in concealed object scenarios. The annotation process, facilitated by Roboflow, not only organized the dataset but also revealed instances with null annotations, adding a layer of complexity to the dataset. This

meticulous dataset preparation serves as a critical foundation for the subsequent experiments and evaluations in the realm of terahertz-based concealed object detection.

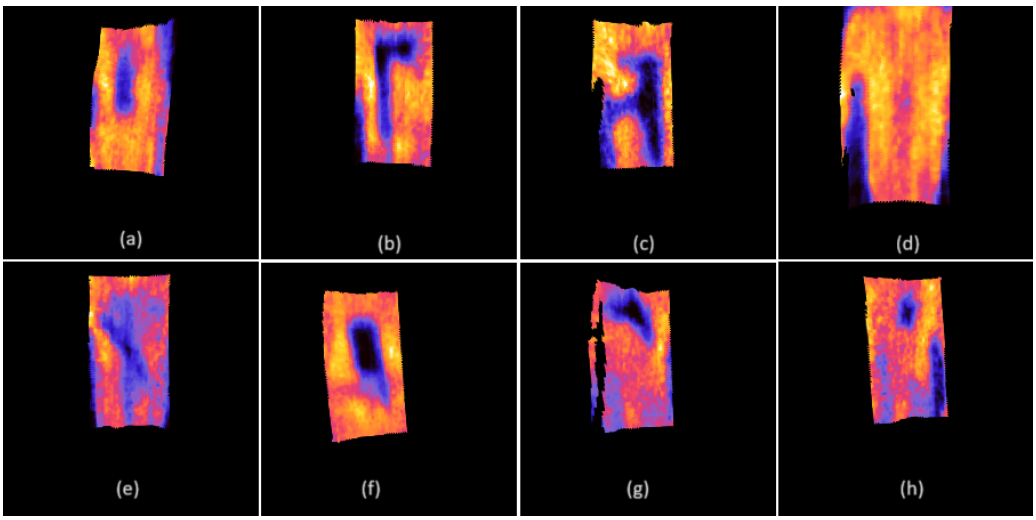

Figure 4.1: Dataset overview

Object Descriptions: (a) bottle (b) axe (c) AK (d) null (no object) (e) knife (f) meatKnife (g) pistol (h) cigaretteBox

The dataset we used is divide into 2301 frames for training, 256 frames for valid, and 256 images for testing in an 80-10-10 ratio. The dataset was prepared by resizing to 640x640 and applying data augumentations on a dataset containing 1274 total images.

Table 4.2: Data Augumentations

Transformation	**Range**
Flip	Horizontal and Vertical
Rotation	Between $-30°$ and $+30°$

4.2 Model Selection and Implementation

In this work, we employ the YOLOv5m model, well-suited for real-time applications, to analyze a terahertz video dataset. The model's architecture, featuring components like the stem, backbone (CSPDarkNet53), neck, and head, enables efficient object detection. The CSPDarkNet53 backbone

plays a crucial role in extracting high-level features for precise object identification within the dataset.

4.2.1 YOLOv5m

The YOLOv5m model, renowned for its enhanced flexibility, stands at the forefront of real-time applications dedicated to concealed object detection. Fig. 3.1 illustrates the comprehensive architecture of YOLOv5m, intricately divided into four integral components: the stem, backbone, neck, and head. At the model's initiation, a set of images, presented in tensor form, serves as the foundational input. The stem module takes center stage as the primary processing unit, undertaking fundamental preprocessing and feature extraction tasks. This crucial step transforms raw input—be it an image or a video frame—into a format conducive to subsequent processing. Typically comprising convolutional layers, batch normalization, and activation functions like ReLU, the stem module sets the groundwork for downstream operations.

4.2.2 Structure of YOLOv5m

The yolov5m architecture is mainly divided into backbone, neck and head section. The details of the structure is given below:

4.2.3 Backbone of YOLOv5m : CSPDarknet

The backbone of YOLOv5m, known as CSPDarkNet53, serves as a critical component for robust feature extraction. Designed as a Convolutional Neural Network (CNN) architecture, CSPDarkNet53 consists of 53 sequentially arranged convolutional layers. What sets it apart is the incorporation of residual connections and bottleneck layers, enhancing its ability to capture intricate patterns and features in input data. A notable innovation within CSPDarkNet53 is the introduction of the "cross-stage partial connection" mechanism. This innovation optimizes information flow across different stages of the network, contributing to heightened performance in object feature extraction. Furthermore, the architecture adopts a "channel-splitting" technique, amplifying the processing of feature maps and ultimately leading to more effective extraction of object-related features.

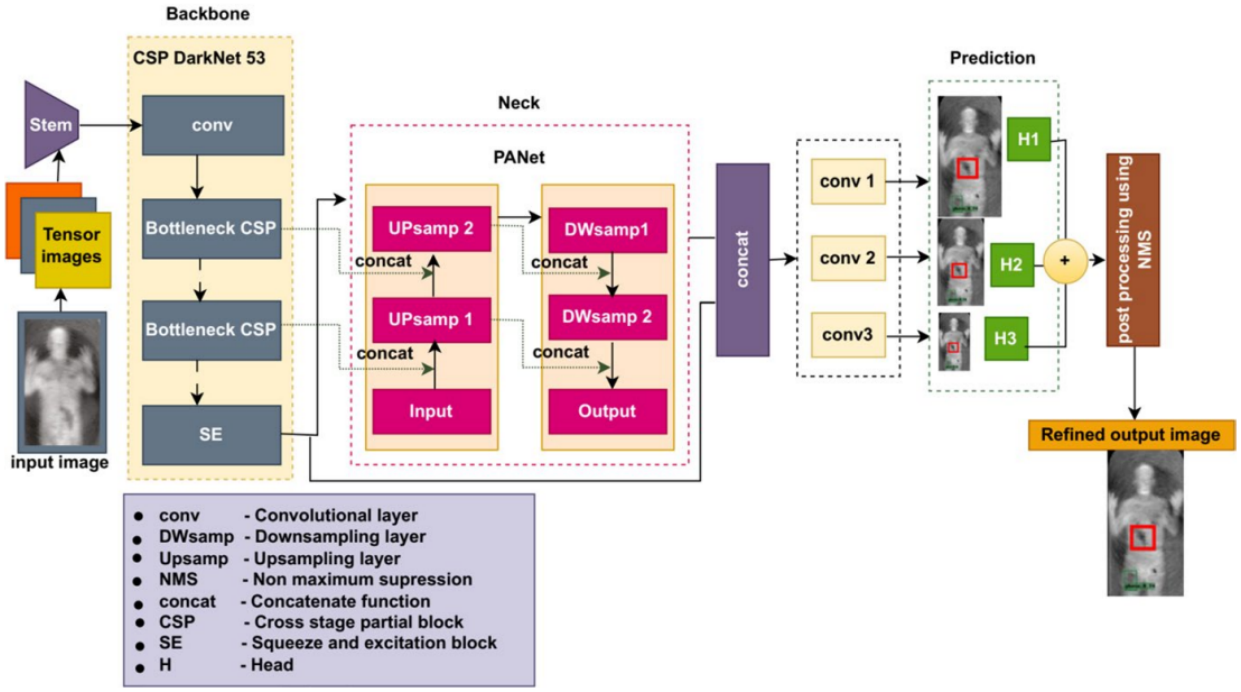

Figure 4.2: Architecture of YOLOv5m

4.2.4 Neck of YOLOv5m : Path Aggregation Network (PANet)

The neck of the YOLOv5m model incorporates the Path Aggregation Network (PANet), which concatenates features from the backbone and PANet output. These combined features undergo convolution layers to reduce computational complexity. PANet includes convolutional layers, BatchNormalization, and ReLU activation, with upsampling and downsampling. It enhances accuracy by preserving fine details and spatial structure through operations like bilinear interpolation and max pooling. Batch Normalization ensures stability and faster training, addressing the vanishing gradient problem, while ReLU activation removes negative values. This configuration contributes to improved accuracy in concealed object detection.

4.2.5 Head of YOLOv5m

The YOLOv5m model's head is pivotal for predicting various aspects based on refined features obtained from the backbone and neck. Comprising three detection heads, this component ensures comprehensive detection capabilities.

1. Firstly, bounding box predictions are made, representing each object with a set of four coordinates relative to the grid cell.

2. Secondly, objectness scores indicate the likelihood of an object's presence in each grid cell, with a scalar value between 0 and 1. A score of 0 signifies no object, while 1 indicates object presence.

3. Lastly, class probabilities are predicted, expressing the likelihood of the detected object belonging to a specific predefined class.

The multi-scale prediction strategy adapts to objects of diverse sizes and resolutions, enhancing the model's versatility in concealed object detection scenarios.

4.2.6 Loss Function

The loss function is a crucial element during the training of YOLOv5m. It consists of three components: Objectness loss (Lo), Classification loss (Lc), and Location loss (LlR). Lo focuses on predicting objectness accurately, Lc on classifying objects correctly, and LlR on refining the spatial location of objects. The loss function equations are:

$$L = L_o + L_c + L_{IR} \tag{4.1}$$

$$Lo = -\sum_{m=0}^{g^2}\sum_{n=0}^{b} L_{m,n}^{o}\left[\hat{c}_m \log(c_m) + (1-\hat{c}_m)\log(1-c_m)\right] \tag{4.2}$$

$$-\varphi_{no}\sum_{m=0}^{g^2}\sum_{n=0}^{b} L_{m,n}^{no}\left[\hat{c}_m \log(c_m) + (1-\hat{c}_m)\log(1-c_m)\right]$$

$$L_c = -\sum_{m=0}^{g^2} L_{m,n}^{o}\sum_{n=0}^{b}\left[\hat{p}_m \log(p_m(r)) + (1-\hat{p}_m(r))\log(1-p_m(r))\right] \tag{4.3}$$

$$L_I R = -\sum_{m=0}^{g^2}\sum_{n=0}^{b} L_{m,n}^{o}\left[1 - R + \frac{d^2(B,B^a)}{r^2} + \beta\gamma\right] \tag{4.4}$$

$$\gamma = \frac{4}{\pi^2}\left(\arctan\frac{W^a}{H^a} - \arctan\frac{W}{H}\right)^2 \tag{4.5}$$

$$\beta = \frac{\gamma}{(1-R)-\gamma} \tag{4.6}$$

g - number of grids

b - number of priori boxes in each grid

φ_{no} - weight

$\hat{c}_m$ - anchor frame

$p_m(r)$ - probability of target function with current prior

r - the diagonal distance of the minimum closure region with prediction and the actual box

R - the prediction and boundary

d - Euclidean distance

B, W and H - The center coordinate, width and height of prediction box

B^a, W^a and H^a - The center coordinate, width and height of actual box

β - weight factor

γ - Similarity ratio of length to width

4.3 Experimental Results

4.3.1 Performance Metrices

Various performance matrices are used for evaluation of the YOLOv5m model. They are:

- **P (Precision):** Precision is the ratio of correctly predicted positive observations to the total predicted positives.

$$Precision = \frac{TP}{TP + FP} \tag{4.7}$$

  where TP is true positive and FP is false positive.

- **R (Recall):** Recall is the ratio of correctly predicted positive observations to the all observations

in actual class.

$$Recall = \frac{TP}{TP + FN} \tag{4.8}$$

where TP is true positive and FP is false negative.

- **mAP50:** Mean Average Precision at 50% IoU. It's a common metric for evaluating object detection models. It considers precision and recall across different levels of confidence for the predicted bounding boxes.

- **Mean Average Precision (mAP)** is a comprehensive metric that takes into account the precision-recall performance of the model across various IoU thresholds. The IoU is the overlap between the predicted bounding box and the ground truth bounding box.

4.3.2 Results

The YOLOv5m model has been successfully implemented, yielding promising training results as depicted in Figure 4.1. Notably, the model exhibits commendable performance metrics, with **precision at 0.983, recall at 0.99, mAP50 at 0.995, and an overall mAP of 0.769.** These metrics collectively indicate the model's proficiency in accurately detecting and localizing objects in the given dataset. However, it is worth noting that certain classes, specifically "handGranade," "cigaretteBox," and "tin," demonstrate comparatively lower accuracy. Despite the model's overall success, these specific classes present challenges, possibly due to factors such as object complexity, variability in appearance, or limited training data for these classes. Addressing these challenges may involve further fine-tuning the model, augmenting the dataset with diverse instances of these classes, or exploring advanced techniques to enhance the model's robustness and accuracy for these specific object categories. Continuous refinement and optimization efforts targeted at the identified classes can contribute to achieving even higher levels of accuracy and reliability across the entire spectrum of object detection tasks.

```
YOLOv5m summary: 212 layers, 20937795 parameters, 0 gradients, 48.1 GFLOPs
                 Class     Images  Instances          P          R      mAP50
                   all        256        246      0.983      0.999      0.995      0.769
               A4paper        256          6      0.978          1      0.995      0.874
                    AK        256         11      0.986          1      0.995      0.879
          AK_noMagazine        256          5      0.971          1      0.995      0.912
                   M16        256          6      0.973          1      0.995      0.846
                   Tin        256         10      0.966          1      0.995      0.634
                   axe        256         11      0.986          1      0.995      0.868
           beltholster        256          8      0.983          1      0.995      0.815
                bottle        256         23      0.998          1      0.995      0.767
           candyboxLid        256         10       0.98          1      0.995      0.794
         cigaretteBox        256         10      0.987          1      0.995      0.635
                fomka        256          7      0.981          1      0.995      0.699
             glassjar        256          8      0.986          1      0.995      0.683
       hammerAndSickle        256         10      0.987          1      0.995      0.812
           handGranade        256         22      0.994          1      0.995      0.591
                 knife        256          4      0.961          1      0.995      0.796
             meatKnife        256         13      0.991          1      0.995       0.73
            phoneNokia        256          8       0.99          1      0.995      0.742
           phoneXiaomi        256          3      0.959          1      0.995      0.752
                pistol        256         39      0.999          1      0.995      0.721
           saucepanLid        256          6       0.97          1      0.995      0.929
       shoulderholster        256         18      0.991          1      0.995      0.822
               usbDisk        256          8          1      0.987      0.995      0.622
```

Figure 4.3: YOLOv5m model summary

Chapter 5

Conclusions and Future Scope

5.1 Conclusion

Utilizing the YOLOv5m model on terahertz images resulted in the successful training of a model, yielding satisfactory outcomes with notable accuracy. The results of the model training were precision at 0.983, recall at 0.99, mAP50 at 0.995, and an overall mAP of 0.769.

However, the observed high accuracy raises concerns about the potential presence of overfitting. This issue is particularly noteworthy given the limited dataset, consisting primarily of singular videos for most classes. Effectively addressing the challenge of overfitting becomes a critical consideration in ensuring the robustness and generalization capabilities of the trained model. The singular nature of the dataset, containing only one video for each class, emphasizes the need for careful regularization techniques and model evaluation to mitigate the risk of overfitting and enhance the model's overall performance on unseen data.

5.2 Future Scope

In the scope of this research, YOLOv5m was employed as the chosen model for object detection. However, with the subsequent release of newer versions such as YOLOv6, YOLOv7, and YOLOv8, there exists the potential for improved accuracy through hyperparameter tuning on these advanced models. The decision to use YOLOv5m may have been influenced by time constraints, leading to the

utilization of only a subset of the dataset. It is plausible that employing the complete dataset could yield enhanced accuracy and performance.

Furthermore, the research highlights a notable limitation in the availability of terahertz datasets, introducing a significant research gap in this domain. The development of comprehensive terahertz datasets is deemed essential for further advancements in research within this field. Bridging this gap will not only contribute to refining the accuracy of current models but also foster innovation and exploration in terahertz-related studies, offering new insights and opportunities for future research endeavors.

Bibliography

[1] Brad Dwyer and Joseph Nelson. Roboflow. Roboflow (Version 1.0) [Software], 2022. Computer Vision.

[2] J Jayachitra, K Suganya Devi, SV Manisekaran, and Satish Kumar Satti. Terahertz video-based hidden object detection using yolov5m and mutation-enabled salp swarm algorithm for enhanced accuracy and faster recognition. *The Journal of Supercomputing*, pages 1–26, 2023.

[3] M Kowalski. Hidden object detection and recognition in passive terahertz and mid-wavelength infrared. *Journal of Infrared, Millimeter, and Terahertz Waves*, 40(11-12):1074–1091, 2019.

[4] Alexei A Morozov and Olga S Sushkova. Development of a publicly available terahertz video dataset and a software platform for experimenting with the intelligent terahertz visual surveillance. In *Proceedings of International Conference on Frontiers in Computing and Systems: COMSYS 2020*, pages 105–113. Springer, 2021.

[5] Chen Wang, Jun Shi, Zenan Zhou, Liang Li, Yuanyuan Zhou, and Xiaqing Yang. Concealed object detection for millimeter-wave images with normalized accumulation map. *IEEE Sensors Journal*, 21(5):6468–6475, 2020.

[6] Xinlin Wang, Shuiping Gou, Jichao Li, Yinghai Zhao, Zhen Liu, Changzhe Jiao, and Shasha Mao. Self-paced feature attention fusion network for concealed object detection in millimeter-wave image. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(1):224–239, 2021.

[7] Xi Yang, Tan Wu, Lei Zhang, Dong Yang, Nannan Wang, Bin Song, and Xinbo Gao. Cnn with spatio-temporal information for fast suspicious object detection and recognition in thz security images. *Signal Processing*, 160:202–214, 2019.