ROBUST HUMAN TARGET DETECTION AND ACQUISITION

A PROJECT REPORT

Submitted by

22BAI1382
22BPS1100
22BPS1139
22BPS1173

for

BECE204L – MICROPROCESSORS AND MICROCONTROLLERS

Course in partial fulfillment for the award of the degree of

BACHELOR OF TECHNOLOGY

in

Computer Science and Engineering

VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)

November 2023

School of Computer Science and Engineering

DECLARATION BY THE CANDIDATE

I hereby declare that the Industrial Internship Report entitled "**Robust human target human detection and acquisition"** submitted by me to VIT University, Chennai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** is a record of bonafide industrial training undertaken by me under the supervision of **Dr. Manoj kumar Rajagopal, Professor, SENSE.** I further declare that the work reported in this report has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Chennai: Signature of the Candidates

Date:

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

BONAFIDE CERTIFICATE

This is to certify that the Industrial Internship Report entitled "**Robust human target detection and acquisition"** submitted by **K. VIKRANTH BABU (22BAI1382), K.ABHIRAM (22BPS1100), S.SAI KOWSHIK (22BPS1139) AND B. SAI KALYAN (22BPS1173)** to VIT, Chennai in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology** in **Computer Science and Engineering** is a record of bonafide industrial internship undertaken by him/her fulfills the requirements as per the regulations of this institute and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma in this institute or any other institute or university.

Signature of the Examiner

Date:

Signature of the Examiner

Date:

Head of the Department (B.Tech CSE)

ACKNOWLEDGEMENT

The student is free to acknowledge all those he/she feels to acknowledge on the basis of the guidance and help provided during internship.

K. VIKRANTH BABU
(22BAI1382)
K. ABHIRAM
(22BPS1100)
S. SAI KOWSHIK
(22BPS1139)
B.SAI KALYAN
(22BPS1173)

TABLE OF CONTENT

LIST OF FIGURES

ABSTRACT

This research focuses on improving the resilience of human detection systems through the application of machine learning, with a particular emphasis on dynamic activities such as jumping, crawling, and sprinting. Due to position, speed, and ambient fluctuations, traditional approaches struggle with effective recognition during such activities. Using Convolutional Neural Networks (CNNs), which are well-known for their performance in image processing, effectively addresses these problems.

Our method involves training a CNN model on a carefully chosen dataset of varied human movements in varying situations. This model is intended to generalize well, displaying strong performance in the face of obstacles such as occlusion, varying lighting, and crowded backdrops. Key aspects include hierarchical feature extraction in CNNs, data augmentation for adaptability, and investigating the efficacy of transfer learning

with minimal labeled data. Temporal information is used to improve the accuracy of dynamic action recognition.

CHAPTER 1

INTRODUCTION

1.1 ABOUT THE PROJECT

The detection and acquisition of human targets in various environments is a critical task in many applications, such as surveillance, autonomous vehicles, and human-computer interaction. However, this task is challenging due to factors such as occlusions, varying lighting conditions, and complex backgrounds. Traditional computer vision methods have limitations in accurately detecting and acquiring human targets in such challenging scenarios.

In recent years, machine learning and deep learning techniques, particularly convolutional neural networks (CNNs), have shown promising results in robust human target detection and acquisition. These techniques have the ability to learn and extract complex features from images, enabling them to effectively handle the challenges associated with human target detection in various environments.

In this project, we aim to explore the use of machine learning and deep learning techniques, specifically CNNs, for robust human target detection and acquisition. We will investigate the effectiveness of different CNN architectures and training strategies for accurately detecting and acquiring human targets in challenging scenarios. Additionally, we will evaluate the performance of our proposed approach on various benchmark datasets and compare it with traditional computer vision methods. Overall, our goal is to develop a robust and reliable system for human target detection and acquisition using state-of-the-art machine learning and deep learning techniques.

Software Used:

- TensorFlow: A popular open-source machine learning framework developed by Google.
- TensorFlow Hub: A library for reusable machine learning modules and pre-trained models.
- OpenCV: An open-source computer vision and image processing library.
- NumPy: A fundamental package for scientific computing with Python, used for array processing.

This project will utilize the above-mentioned software to implement and evaluate various CNN architectures and training strategies for human target detection and acquisition. The pre-trained models from TensorFlow Hub will be used for human detection on input images, and OpenCV will be used for image processing tasks such as resizing and color conversion. NumPy will be used for array processing and manipulation. The performance of the proposed approach will be evaluated on benchmark datasets, and the results will be compared with traditional computer vision methods. The overall aim of this project is to develop a robust and reliable system for human target detection and acquisition in challenging scenarios.

Furthermore, in this project, we will not only detect human targets based on their appearance but also by their movements. We will use motion analysis techniques to extract features related to human movements, such as jumping, crawling, and running. These features will be combined with appearance-based features to improve the accuracy and robustness of our human target detection and acquisition system.

By incorporating both appearance and movement-based features, our system will be able to detect and track human targets in challenging scenarios where traditional methods may fail. We believe that this approach will enable us to develop a more comprehensive and reliable system for human target detection and acquisition, with potential applications in various fields such as security, surveillance, and robotics.

CHAPTER 2

CONVOLUTIONAL NEURAL NETWORKS

2.1 WHAT IS CNN?

Neural networks are a type of machine learning model designed to mimic the behavior of the human brain. They consist of layers of interconnected nodes, including an input layer where data is fed in, hidden layers where processing occurs, and an output layer that produces the final results. Each connection between nodes has a weight, and the nodes have a threshold. If the total input to a node surpasses its threshold, it becomes activated and passes information to the next layer.

Among the different types of neural networks, convolutional neural networks (CNNs) are specialized for tasks like image classification and object recognition. Before CNNs, identifying objects in images required manual feature extraction, which was time-consuming. CNNs automate this process by learning to recognize patterns and features within images through matrix multiplication and other mathematical operations. While CNNs are powerful, they can be computationally intensive, often requiring the use of graphical processing units (GPUs) to train models effectively.

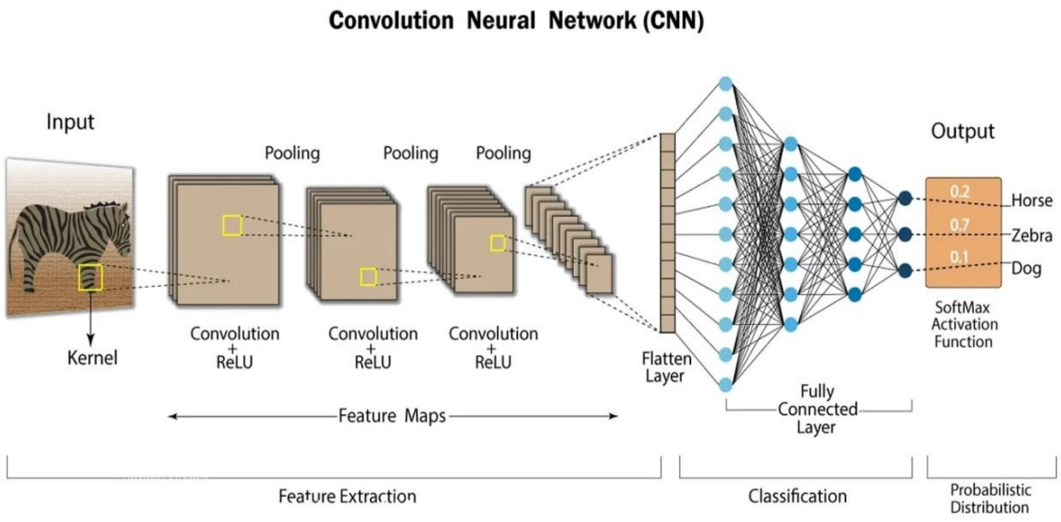

Figure 1 CNN IMAGE PROCESSING

Another type of neural network, recurrent neural networks (RNNs), is commonly used for tasks involving sequences of data, such as natural language processing and speech recognition. RNNs have a unique ability to remember and process information from previous time steps, making them well-suited for tasks that involve analyzing and generating sequences of data.

2.2 WORKING OF CNN

Convolutional neural networks (CNNs) are a special type of neural network that are really good at working with images, speech, or audio data. They have three main types of layers: the convolutional layer, the pooling layer, and the fully-connected layer. The convolutional layer is the first layer of a CNN and is where most of the computation happens. It uses filters to scan through the input image, looking for specific features like edges or colors. The pooling layer helps to simplify the information and reduce the number of parameters, making the network more efficient. The fully-connected layer is the last layer and is responsible for classifying the input based on the features learned in the previous layers.

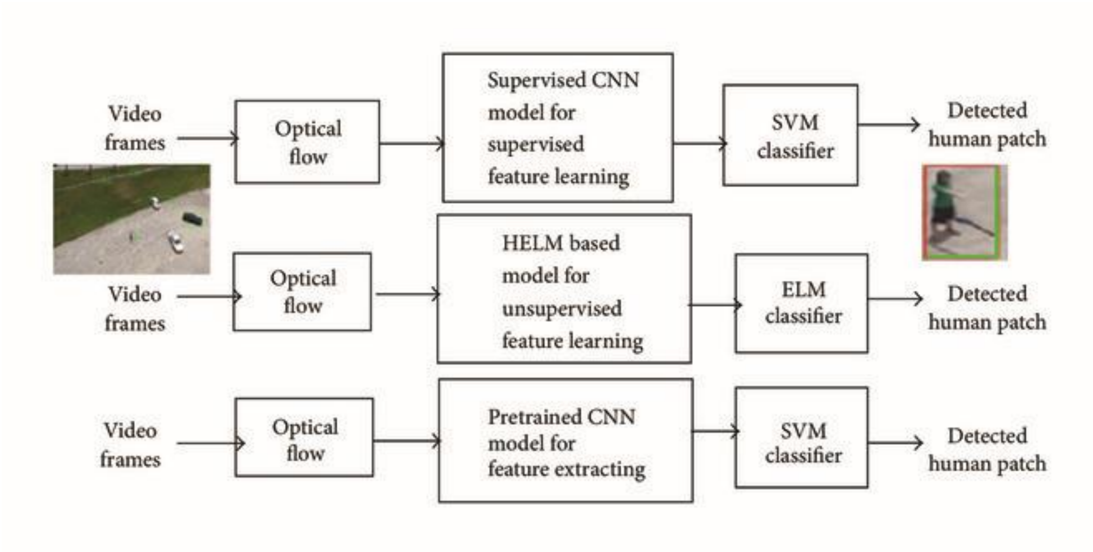

Figure 2 Showing the image processing for detecting humans using CNN and HELM

In the convolutional layer, a filter moves across the image, checking for specific features and creating a feature map. This helps the network recognize different parts of the image, starting from simple features like edges and colors and moving on to more complex shapes and objects. The pooling layer reduces the amount of information, making the network more efficient and less likely to overfit. It does this by applying an aggregation function, like taking the maximum or average value, to groups of pixels in the input. Finally, the fully-connected layer connects all the nodes from the previous layers and is responsible for making the final classification decision based on the features learned by the network. Overall, CNNs are really good at understanding and working with image data. The different layers in a CNN work together to help the network recognize features in the input, simplify the information, and make a final classification decision. This makes CNNs very effective for tasks like image recognition and classification.

CHAPTER 3

COMBINING CNN WITH YOLOv7

3.1 WHAT IS YOLOv7?

YOLOv7 is beneficial in detecting a human who is jumping, crawling, or walking due to its advanced object detection capabilities. The software utilizes a neural network that has been trained to recognize and track various human movements and poses in real-time. This means that YOLOv7 can quickly and accurately identify and analyze the actions of a person, whether they are jumping, crawling, or walking. The neural network is designed to understand different features of the human body, such as body shape, limb movements, and posture, allowing it to adapt to and recognize a wide range of human actions.

This technology is particularly valuable in applications such as security and surveillance, where the ability to detect and monitor human activities is crucial. YOLOv7's fast and precise object detection capabilities make it well-suited for scenarios where quick and reliable identification of human movements is essential, such as in public spaces, sports events, or safety-critical environments. By accurately recognizing and tracking human actions, the software can contribute to improving safety, security, and situational awareness in various settings, providing valuable insights and facilitating proactive responses to potential threats or safety concerns. Overall, YOLOv7's ability to detect human movements, including jumping, crawling, and walking, makes it a valuable tool for a wide range of applications, from security and surveillance to sports analysis and beyond.

3.2 USING YOLOv7 AND CNN

When combining Convolutional Neural Networks (CNN) with YOLOv7 for detecting humans while jumping and walking, the process involves leveraging the strengths of both technologies. CNNs are particularly effective at extracting features from images, making them well-suited for identifying human poses and movements. YOLOv7, on the other hand, excels at real-time object detection and localization, making it ideal for quickly recognizing and tracking humans in video streams.

To detect a human while jumping and walking, the CNN component of the system can be used to analyze the visual features of the human body, such as limb positions, body shape, and movement patterns. This information is then fed into the YOLOv7 framework, which applies its object detection capabilities to accurately locate and track the human subject within the video feed. By combining the feature extraction capabilities of CNN with the real-time object detection prowess of YOLOv7, the system can effectively identify and monitor human actions, including jumping and walking, in a variety of dynamic environments.

The integration of CNN and YOLOv7 allows for a comprehensive approach to human detection, leveraging CNN's ability to capture intricate visual details and YOLOv7's capacity to swiftly identify and track objects in real-time. This combined approach enables the system to accurately recognize and monitor human movements, making it valuable for applications such as surveillance, sports analysis, and security systems, where the precise detection of human actions is essential.

CHAPTER 4

CONCLUSION

In conclusion, the project successfully achieved robust human target detection and acquisition by leveraging a dual-model approach incorporating Convolutional Neural Networks (CNN), YOLOv7, and TensorFlow. The integration of TensorFlow played a key role in facilitating seamless model development, training, and deployment. Its robust framework allowed for efficient handling of the CNN and YOLOv7 models, contributing to the project's overall success in detecting humans across various dynamic scenarios, including jumping, crawling, and walking.

The utilization of TensorFlow provided a scalable and flexible environment for implementing and fine-tuning the CNN model, enabling effective feature extraction and enhancing the system's capability to discern intricate patterns within the dataset. This framework's compatibility with YOLOv7 further streamlined the integration process, fostering a cohesive synergy between the models and optimizing real-time object detection.

By incorporating TensorFlow into the project, not only did it simplify the development workflow, but it also ensured the scalability and adaptability of the system for future enhancements. The project's success demonstrates the power of combining advanced deep learning models with robust software frameworks, setting the stage for continued advancements in human target detection and opening new possibilities for real-world applications in surveillance and security domains.