COMPUTER VISION GUIDED LIFE DETECTION DURING FLOODS

A Project Report

submitted to the APJ Abdul Kalam Technological University

in partial fulfillment of the requirements for the award of degree

Bachelor of Technology

in

Computer Science and Engineering

by

Adithya Vinod(TCR20CS009)

Annu Jo(TCR20CS018)

Nandana K Ajay(TCR20CS048)

Nandana S Kartha(TCR20CS049)

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GOVERNMENT ENGINEERING COLLEGE, THRISSUR

KERALA

December 2023

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

GOVERNMENT ENGINEERING COLLEGE

THRISSUR - 680009

CERTIFICATE

This is to certify that the report entitled **COMPUTER VISION GUIDED LIFE DETECTION DURING FLOODS** submitted by **Adithya Vinod** (TCR20CS009), **Annu Jo** (TCR20CS018), **Nandana K Ajay** (TCR20CS048) & **Nandana S Kartha** (TCR20CS049) to the APJ Abdul Kalam Technological University in partial fulfillment of the B.Tech. degree in Computer Science and Engineering is a bonafide record of the project work carried out by him/her under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Panchami V U
(Project Guide)
Assistant Professor
Department of CSE
GEC Thrissur

Princy Ann Thomas
(Project Coordinator)
Assistant Professor
Department of CSE
GEC Thrissur

George Mathew
(Project Coordinator)
Assistant Professor
Department of CSE
GEC Thrissur

Dr. Shibily Joseph
Head of the Department
Associate Professor
Department of CSE
GEC Thrissur

DECLARATION

We hereby declare that the project report **COMPUTER VISION GUIDED LIFE DETECTION DURING FLOODS** , submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by us under supervision of **Panchami V U**.

This submission represents our ideas in our own words and where ideas or words of others have been included, we have adequately and accurately cited and referenced the original sources.

We also declare that we have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in our submission. We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Adithya Vinod

Annu Jo

Nandana K Ajay

Nandana S Kartha

Thrissur

December 4, 2023

Abstract

Floods pose significant threats to human lives and property, demanding efficient and timely response mechanisms. Floods in India, have been a recurring natural disaster, causing significant damage to infrastructure and human lives. A computer vision-guided life detection system can be highly relevant in the context of managing and responding to such floods for several reasons such as rapid and efficient search and rescue operations, real-time data analysis, nighttime and low visibility operations, post-disaster assessment, etc.

The proposed system employs cameras and computer vision algorithms to capture, process, and analyse visual data from flood-affected areas. These algorithms detect and track objects, distinguishing between living beings and non-living objects. Machine learning models are then used to identify specific human body shapes, movements, and gestures, enhancing the system's ability to identify signs of life. The output of this system is communicated to emergency responders, providing real-time information about the presence and locations of potential victims in floodwaters. The computer vision-guided life detection system during floods represents a promising advancement in disaster response technology. By augmenting traditional methods with artificial intelligence and visual analysis, it has the potential to save lives and minimize the impact of floods on communities. Future developments will focus on refining accuracy, real-time processing capabilities, and seamless integration into existing emergency response frameworks.

Overall, our computer vision-guided Life detection web application provides an accessible and affordable solution.

Acknowledgement

We take this opportunity to expressoury deepest sense of gratitude and sincere thanks to everyone who helped us to complete this work successfully. We express our sincere thanks to **Dr. Shibily Joseph**, Head of Department, Computer Science and Engineering, Government Engineering College, Thrissur for providing us with all the necessary facilities and support.

We would like to express our sincere gratitude to our project coordinators **Princy Ann Thomas**, Assistant Professor and **George Mathew**, Assistant Professor, Department of Computer Science and Engineering, Government Engineering College, Thrissur for the support and co-operation.

We would like to place on record our sincere gratitude to our project guide **Panchami V U**, Assistant Professor, Computer Science and Engineering, Government Engineering College, Thrissur for the guidance and mentorship throughout this work.

Finally we thank our families, and friends who contributed to the successful fulfilment of this project work.

Adithya Vinod

Annu Jo

Nandana K Ajay

Nandana S Kartha

Contents

List of Figures

List of Tables

ABBREVIATIONS

CNN	Convolutional Neural Network
ResNet	Residual Neural Network
YOLO	You Only Look Once
HTML	Hyper Text Markup Language
CSS	Cascading Style Sheets
Wi-Fi	Wireless Fidelity
UAV	Unmanned Aerial Vehicle
DNN	Deep Neural Network
SiLU	Sigmoid Linear Unit
FReLU	Flexible Rectified Linear Unit
SSD	Solid State Drive
RCNN	Region Based Convolutional Neural Networks
GPU	Graphics Processing Unit
LSTM	Long Short-Term Memory
LRCN	Long-term Recurrent Convolutional Network
RNN	Recurrent Convolutional Network

Chapter 1

Introduction

1.1 Topic Introduction

In the face of recurring natural disasters, particularly the devastating floods that frequently afflict regions like India, the demand for innovative and efficient technologies in disaster response has become increasingly evident. This project introduces a computer vision-guided life detection system designed to revolutionize flood management and response strategies. Leveraging cutting-edge technologies such as cameras, computer vision algorithms, and machine learning models, the system aims to swiftly and accurately identify living beings in flood-affected areas. By distinguishing between living and non-living objects and employing artificial intelligence to recognize specific human body shapes, movements, and gestures, the system provides real-time information to emergency responders, facilitating rapid search and rescue operations. The integration of advanced visual analysis into traditional disaster response methods marks a promising step towards minimizing the impact of floods on communities and saving lives. This project report delves into the detailed development, functionality, and future enhancements of the computer vision-guided life detection system, emphasizing its significance in the broader context of flood management and disaster resilience.

1.2 Problem Statement

Floods are natural disasters that have the potential to cause extensive damage to property and, more critically, to put human lives at risk. Traditional methods of flood response and search and rescue operations rely heavily on human resources, often facing significant challenges, such as limited visibility, treacherous conditions, and the need for timely and accurate decision-making. This project aims to address the critical issues in flood response.

1.3 Objectives

Develop a real-time flood monitoring system: Create a system that can monitor flood conditions and water levels using cameras and drones.

Implement computer vision algorithms: Develop and implement computer vision algorithms to analyze live camera feeds to identify individuals in distress during a flood.

Person detection and tracking: Design algorithms that can accurately detect and track the presence and movement of people in flooded areas.

Object recognition: Enable the system to recognize and distinguish between people, debris, and other objects in the floodwaters.

SOS signal recognition: Implement technology to detect visual and audible distress signals, such as waving or shouting for help.

Disaster response coordination: Promote better coordination among emergency response teams by providing them with accurate information to enhance their rescue efforts.

1.4 Novelty of Idea and Implementation Steps

Novelty of Idea:

Our system introduces a groundbreaking approach to flood management by harnessing the power of computer vision and machine learning. Unlike traditional methods, our system dynamically identifies and tracks living beings in flood-affected areas in real-time. By focusing on recognizing specific human features and movements, it enhances accuracy, providing instant, valuable information for emergency responders. This fusion of artificial intelligence and visual analysis stands as an innovative leap forward, promising a more effective and timely response to flood-related emergencies.

Implementation Steps:

System Design: Begin by designing the architecture of the computer vision-guided life detection system. Define the components, their interactions, and the overall flow of data within the system.

Data Collection: Set up cameras and drones in flood-affected areas to capture real-time images and video feeds. Accumulate a diverse dataset for training and testing machine learning models.

Algorithm Development: Develop computer vision algorithms capable of distinguishing between living beings and non-living objects in the captured data. Implement machine learning models, particularly focusing on identifying human body shapes, movements, and gestures.

Integration of Models: Integrate the developed algorithms and machine learning models into the system's backend, ensuring seamless communication and data processing.

User Interface Design: Design user interfaces using HTML, CSS, and JavaScript to facilitate real-time monitoring and interaction for users. Implement features for object differentiation and alert management.

Backend Implementation: Utilize the Flask web framework for the system's backend, managing user accounts, handling alerts, and controlling overall system operation.

Communication Interfaces: Implement communication interfaces, such as Wi-Fi or cellular networks, for transmitting real-time alerts and notifications to emergency responders.

Testing and Validation: Conduct thorough testing to ensure the accuracy and reliability of the object detection models and the overall system. Address any issues or inaccuracies through iterative refinement.

Privacy and Security Measures: Implement robust privacy and security measures to handle sensitive data, ensuring compliance with privacy regulations. Encrypt data transmission and establish user authentication mechanisms.

Deployment: Deploy the computer vision-guided life detection system in flood-prone areas, integrating it into existing emergency response frameworks.

Monitoring and Maintenance: Establish a system for continuous monitoring and maintenance to address any issues, update algorithms/models, and ensure the system's optimal performance over time.

1.5 Societal and Industrial Relevance

Our project addresses a pressing issue – the impact of floods on lives and infrastructure, especially in places like India. From an industry perspective, it modernizes disaster response, offering faster and more accurate search and rescue operations. This aligns with the increasing demand for advanced technologies in disaster management. On a societal level, the system directly tackles a crucial problem by swiftly identifying people in floodwaters, ensuring timely and precise rescue efforts. This technology has the potential to significantly reduce casualties and boost the effectiveness of emergency responders. By combining cutting-edge technology with real-world challenges, our project becomes a valuable tool with broad benefits for both the industry and communities affected by natural disasters.

Chapter 2

Literature Review

2.1 Topic of review-Section Wise Review

2.1.1 Human Detection Based Yolo Backbone-Transformer in UAV

This study presents a new method for human detection in UAVs using Yolo backbones transformer. The proposed framework utilizes backbones YoloV8s, SC3T (Based Transformer), with RGB inputs to accurately perceive human detection.The effectiveness and efficiency of human detection utilizing YOLO or any object detection algorithm may be influenced by several factors, such as the implementation, the dataset used for training, the computational resources available, and the prevailing environmental conditions. The superior performance of our Deep Neural Network (DNN) can provide context awareness to UAVs. These algorithms can be trained on large datasets of human images, enabling them to learn features and patterns associated with human appearance, motion, and behavior. This work highlights the potential of the Yolo backbones transformer for enhancing human detection in UAVs, demonstrating its superiority over conventional methods. Overall, the proposed framework can pave the way for future research in object detection. [1]

2.1.2 An Improved Swimming Pool Drowning Detection Method Based on YOLOv8

This paper introduces an enhanced drowning detection model based on YOLOv8, addressing spatial perception and accuracy. Retaining YOLOv8's main structure, the model incorporates a lightweight coordinate attention (CA) mechanism and the FReLU activation function for improved detection. CA enhances direction and position perception on each channel, surpassing spatial or channel attention mechanisms. FReLU replaces SiLU, enhancing spatial perceptibility without significant spatial overhead. Tested on a swimmer dataset, the proposed method achieves a 91.78% mean average precision, a 1.63% improvement over YOLOv8. The paper emphasizes real-time, accurate computer vision drowning detection, positioning YOLOv8 as a performant choice. Contributions include the addition of a lightweight CA mechanism and the replacement of SiLU with FReLU, enhancing model accuracy and spatial perception. Experimental results validate the improvements on the constructed dataset. The paper concludes with a concise summary of the technical advancements. [2]

2.1.3 YOLO-ResNet: A New Model for Rebar Detection

The paper proposes a novel object detection model based on YOLOv3 and ResNet-18 for detecting steel bars in construction images. The model uses CBAM attention modules, new anchor boxes, and FocalLoss to improve the performance and efficiency of rebar detection. The paper uses a dataset of 450 images of rebar with different angles, lighting, and density. The paper uses four metrics to evaluate the model: TrainTime, BestLoss, mAP, and Accuracy. The paper compares the proposed model with SSD, Faster RCNN, and YOLOv3 on these metrics. The paper shows that the proposed model achieves better results than the other models on all metrics, with faster training time, lower GPU memory usage, smaller model weight size, and higher detection accuracy. The paper also shows some examples of the detection results on the test dataset. [3]

2.1.4 Identification of Drowning Victims in Freshwater Bodies

The paper addresses the problem of detecting and rescuing drowning victims in freshwater ecosystems, which is the third leading cause of unintentional injury deaths worldwide. The paper proposes a portable system that can calculate the drift distance of a victim based on water current and body parameters, and detect the victim using a combination of ultrasonic sensor and underwater camera with YOLOv3 object detection algorithm. The paper describes the design, implementation and testing of the proposed system, which consists of four modules: data collection and training, drift prediction, image classification and human detection. The paper reports that the system achieved a mean average precision (mAP) of 98.0% and an average intersection over union (IoU) of 82.10% for object detection, and a detection depth of 5m. The paper also shows the accuracy of drift prediction for different water currents and body parameters. [4]

2.1.5 Multiple Human Objects Tracking in Long Range Infrared Videos

The paper presents a novel approach to detect, track, and classify human objects in low quality and long range infrared videos using deep learning algorithms directly on compressive measurements, without reconstructing the original frames. The paper applies two deep learning models, YOLO for object detection and tracking, and ResNet for human classification, to the compressive measurements generated by pixel subsampling or coded aperture cameras. The paper evaluates the proposed approach using low quality videos in the SENSIAC database, which contains optical and mid-wave infrared (MWIR) videos of human subjects at various ranges up to 1500 meters. The paper reports the performance metrics such as center location error, distance precision, mean average precision, and classification accuracy for different subsampling rates and ranges. [5]

2.1.6 A Stable Deep Learning Model for Recognizing Human Actions in Noisy Videos

The paper develops a deep learning model using ConvLSTM, LRCN and DenseNet architectures for accurately recognizing human actions from noisy videos. The models are trained and tested on the UCF50 and HMDB51 public datasets containing videos of humans performing 50 different actions. The video data is preprocessed via filtering, frame resizing and normalization. 75% of the data is used for training and 25% for testing. The evaluation metric is accuracy. ConvLSTM captures spatial and temporal aspects of frames. LRCN combines CNN for spatial features and LSTM for sequence modelling. DenseNet121+LSTM uses DenseNet for feature extraction from frames and LSTM to capture temporal dependencies. Comprehensive experiments show DenseNet121+LSTM achieves the highest accuracy of 98.8% on UCF50 and 90.8% on HMDB51, outperforming state-of-the-art approaches. Confusion matrix analysis demonstrates high classification accuracy for top classes. DenseNet121+LSTM is recommended as the best model for real-time human activity recognition from noisy videos. LRCN also shows good performance. Extensive experiments demonstrate the effectiveness of the proposed approach. [6]

2.1.7 An Approach to Recognize Human Activities based on ConvLSTM and LRCN

An Approach to Recognize Human Activities based on ConvLSTM and LRCN. The paper develops ConvLSTM and LRCN models for accurate human activity recognition from videos and sensor data. Related works using CNN, LSTM, RNN, ConvLSTM, LRCN along with wearable sensors are reviewed, discussing their merits and limitations. The proposed approach implements a ConvLSTM model to capture spatial and temporal relationships and a LRCN model that combines CNN for feature extraction and LSTM for sequence classification. The UCF50 dataset containing 50 human activity categories from YouTube is used. Data preprocessing like frame extraction, resizing and shuffling is done. 75% of data is used for training and 25% for testing. The ConvLSTM model built using Keras LSTM layers identifies spatial patterns and temporal dependencies, achieving 80.33% accuracy. The LRCN model combines time-distributed CNN layers for feature extraction and LSTM layer for classification, obtaining 92.62% accuracy outperforming other approaches. LRCN predictions are demonstrated on YouTube videos. As the key outcome, LRCN gives the best performance for human activity recognition from videos. The issue of multi-class identification needs to be addressed in future works. [7]

2.1.8 A Deep LSTM Approach for Activity Recognition

The paper focuses on applying a deep Long Short-Term Memory (LSTM) neural network for feature-free classification of human activities using raw accelerometer data. Related works have used feature extraction with traditional machine learning algorithms. Recent approaches explore deep learning methods like CNN-LSTMs for sensor-based activity recognition. The paper collects triaxial accelerometer data for 7 activities from 15 participants. The raw data is normalized and segmented. A deep LSTM model with two hidden layers is implemented for classification. Evaluation uses 10-fold stratified cross validation to assess model performance. The results demonstrate that the deep LSTM model achieves classification accuracy of 91.34% and F1 score of 87.58% which outperforms prior feature-based methods on the dataset. The confusion matrix provides detailed class-wise performance. Activities with fewer samples have relatively lower accuracy. The deep LSTM model processes raw accelerometer data effectively for accurate human activity recognition, eliminating the need for complex feature engineering. [8]

2.1.9 Tracking of Multiple Human Objects Directly in Low Quality Optical Videos

The document presents a study on the use of compressive sensing for low-quality videos, with a focus on target tracking and classification. The authors employ pixel subsampling and coded aperture as methods for compressive measurements. For target tracking, the YOLO algorithm is used, which is trained on cropped targets from slow-paced videos and tested on fast-paced ones. The performance of YOLO is evaluated using four metrics, demonstrating high detection rates and reasonable tracking accuracy for different ranges and missing rates. For target classification, the ResNet deep convolutional neural network is used. It's trained on augmented data from slow-paced videos and tested on fast-paced ones. Despite achieving good classification results for short ranges and low missing rates, the performance drops significantly for longer ranges and higher missing rates. The authors acknowledge the challenges posed by the small size, low quality, and long range of the videos. They suggest further research to improve accuracy, including the use of different compressive sensing devices, improving the data augmentation, and exploring other deep learning models. [9]

2.1.10 Real-Time Human Detection as an Edge Service Enabled by a Lightweight CNN

The document provides a comprehensive review of the application of edge computing in smart surveillance systems, with a particular focus on real-time human object detection and behavior recognition. It introduces a novel Lightweight Convolutional Neural Network (L-CNN) specifically designed for human object detection at the edge. The L-CNN leverages depthwise separable convolution and the single shot multi-box detector (SSD) architectures to reduce computational complexity and memory usage, making it suitable for edge devices like Raspberry PI 3. The L-CNN is trained to focus on human objects as the only class of interest, which narrows down the classifier's searching space and improves the accuracy. The performance of the L-CNN is evaluated using various metrics, including speed, CPU usage, memory usage, and false positive rate. The results show that the L-CNN can process an average of 1.79 and up to 2.06 frames per second (FPS), meeting the real-time requirement for human object detection. It also demonstrates a low false positive rate of 6.6%, which is close to the SSD GoogleNet's 5.3%. However, the document acknowledges the challenges posed by the small size, low quality, and long range of the videos. It suggests further research to improve accuracy, including the use of different compressive sensing devices, improving the data augmentation, and exploring other deep learning models. This literature review provides valuable insights into the current state of edge computing in smart surveillance and highlights potential areas for future research. [10]

Title	Author and Year of Publication	Method	Performance Measure	Summary of Findings
Human Detection Based Yolo Backbone-Transformer in UAVs	Manh-Tuan Do, Manh-Hung Ha, Duc-Chinh Nguyen, Kim Thai, Quang -Huy Do Ba ,2023	YoloV8s, SC3T (Transformer) with RGB inputs	N/A	Proposes a framework for human detection in UAVs using Yolo backbones transformer. Demonstrates effectiveness in perceiving human detection, offering context awareness to UAVs. Highlights potential for future research in object detection.
An Improved Swimming Pool Drowning Detection Method based on YOLOv8	Tianyi He, Xiaodong Ye, Meiling Wang, 2023	YOLOv8 with lightweight CA mechanism	Mean Average Precision (MAP): 91.78%	Introduces an enhanced drowning detection model based on YOLOv8, achieving a 1.63% improvement over YOLOv8. Emphasizes real-time, accurate computer vision drowning detection. Contributions include the addition of a lightweight CA mechanism and FReLU activation.
YOLO-ResNet: A New Model for Rebar Detection	Yaoshun Li, Lizhi Liu, 2021	YOLOv3, ResNet-18 with CBAM, new anchor boxes	Train Time, Best Loss, mAP, Accuracy	Proposes a novel object detection model for rebar detection using YOLOv3 and ResNet-18. Shows superior results compared to SSD, Faster RCNN, and YOLOv3 on various metrics, including faster training time, lower GPU usage, smaller model weight, and higher accuracy.

Figure 2.1: Literature Review(a)

Title	Author and Year of Publication	Method	Performance Measure	Summary of Findings
Identification of Drowning Victims in Freshwater Bodies using Drift Prediction and Image Processing based on Deep Learning	Anjana Unnikrishnan, A T Roshni, P R Anusha, Anju M Vinny, C K Anuraj 2021	YOLOv3 with ultrasonic sensor and underwater camera	Mean Average Precision (mAP): 98.0%	Addresses drowning detection in freshwater ecosystems. Proposes a portable system with YOLOv3 for human detection, achieving high precision. Describes a system consisting of data collection, drift prediction, image classification, and human detection modules.
Multiple Human Objects Tracking in Long Range Infrared Videos	Chiman Kwan, David Gribben, Trac Tran, 2019	YOLO for detection, ResNet for classification	Center Location Error, Distance Precision, mAP, Classification Accuracy	Presents an approach for detecting, tracking, and classifying human objects in long-range infrared videos using YOLO for detection and ResNet for classification. Evaluates performance metrics for different subsampling rates and ranges.
A Stable Deep Learning Model for Recognizing Human Actions in Noisy Videos	Ishwar Chand, Sanjay Kumar, Jatin Popli, Jitesh Kumar, 2023	ConvLSTM, LRCN, DenseNet	Accuracy	Develops deep learning models for recognizing human actions in noisy videos. Achieves high accuracy, with DenseNet121+LSTM outperforming state-of-the-art approaches. Recommends DenseNet121+LSTM for real-time human activity recognition from noisy videos.

Figure 2.2: Literature Review(b)

Title	Author and Year of Publication	Method	Performance Measure	Summary of Findings
An Approach to Recognize Human Activities based on ConvLSTM and LRCN	Shradha Bhatia, Tushar Chauhan, Sumita Gupta, Sapna Gambhir, Jitesh H. Panchal, 2023	ConvLSTM, LRCN	Accuracy	Develops ConvLSTM and LRCN models for accurate human activity recognition. Achieves high accuracy, with LRCN outperforming other approaches. Recommends LRCN for human activity recognition from videos.
A Deep LSTM Approach for Activity Recognition	Selda Güney, Çağatay Berke Erdaş, 2019	Deep LSTM	Classification Accuracy, F1 Score	Focuses on applying a deep LSTM neural network for feature-free classification of human activities using raw accelerometer data. Achieves high classification accuracy, outperforming prior feature-based methods.

Figure 2.3: Literature Review(c)

Title	Author and Year of Publication	Method	Performance Measure	Summary of Findings
Tracking of Multiple Human Objects Directly in Compressive Measurement Domain for Low quality Optical videos	Chiman Kwan, David Gribben, Trac Tran, 2019	YOLO for tracking, ResNet for classification	N/A	Studies compressive sensing for low-quality videos, focusing on target tracking and classification. Uses YOLO for tracking and ResNet for classification. Evaluates performance metrics for different ranges and missing rates.
Real-Time Human Detection as an Edge Service enables by lightweight CNN	Seyed Yahya Nikouei, Yu Chen, Sejun Song, Ronghua Xu, Baek-Young Choi, Timothy R. Faughnan, 2018	Lightweight CNN (L-CNN)	Speed, CPU Usage, Memory Usage, False Positive Rate	Reviews the application of edge computing in smart surveillance. Introduces L-CNN for real-time human object detection. Demonstrates L-CNN's suitability for edge devices. Evaluates performance metrics, including speed, CPU usage, memory usage, and false positive rate.

Figure 2.4: Literature Review(d)

2.2 Conclusions and Gap Analysis

2.2.1 Conclusion

In conclusion, the literature review provides a comprehensive overview of the current state of research in the application of computer vision and deep learning techniques, particularly focusing on the utilization of YOLO and ResNet models in various domains. The reviewed papers demonstrate the versatility of these models in addressing specific challenges, ranging from human detection in UAVs to drowning detection, rebar detection in construction, and human activity recognition.

The YOLO backbones transformer, as showcased in the first paper, presents a promising avenue for accurate human detection in UAVs. Similarly, the enhanced drowning detection method based on YOLOv8, as detailed in the second paper, introduces improvements in spatial perception and accuracy, positioning YOLOv8 as a performant choice. The third paper proposes a novel YOLO-ResNet model for rebar detection, showcasing superior results compared to existing models.

Furthermore, the literature explores applications beyond conventional scenarios, such as the identification of drowning victims in freshwater bodies and tracking multiple human objects in long-range infrared videos. The studies also delve into human activity recognition, addressing the challenges of noisy videos, accelerometer data, and low-quality optical videos.

2.2.2 Gap Analysis

While the reviewed literature provides valuable insights into the capabilities of YOLO and ResNet models in various applications, there are notable gaps that warrant further exploration:

Interdisciplinary Integration: Many studies focus on specific applications within computer vision, but there is a need for more interdisciplinary research that integrates computer vision with other domains such as robotics, environmental science, and healthcare.

Robustness in Challenging Conditions: While the literature covers scenarios like drowning detection in challenging environments, there is room for research that addresses the robustness of these models in extreme conditions, including adverse weather, low lighting, and complex terrains.

Edge Computing for Real-time Processing: Although some papers touch upon real-time processing, there is an opportunity for more research on optimizing deep learning models, like Lightweight CNNs, for edge computing, considering resource-constrained devices.

Standardized Evaluation Metrics: A standardized set of evaluation metrics for different applications could facilitate more consistent comparisons between models and enhance the reliability of their performance assessments.

Addressing these gaps will contribute to the advancement of computer vision applications and deepen our understanding of the capabilities and limitations of YOLO and ResNet models in diverse contexts.

2.3 Summary

2.3.1 Literature Review Summary

The literature review encompasses a diverse range of applications of computer vision and deep learning techniques, predominantly focusing on YOLO and ResNet models. The reviewed papers demonstrate the efficacy of these models in various domains, including human detection in UAVs, drowning detection, rebar detection in construction, and human activity recognition. Notable studies address challenges in real-time processing, spatial perception, and robustness in adverse conditions.

2.3.2 Influence on Project Title

The literature review significantly shapes the project title and objectives by providing a foundation for understanding the existing landscape of computer vision applications. The emphasis on YOLO and ResNet models in diverse scenarios aligns with the project's focus on a "Computer Vision-Guided Life Detection System during Floods." The title reflects the integration of computer vision techniques (YOLO and ResNet) to address a critical issue—efficiently detecting signs of life in flood-affected areas.

2.3.3 Objectives Influenced by the Review

Utilization of YOLO and ResNet Models: The literature review highlights the success of YOLO and ResNet models in various applications. This influences the project's objective to employ these models for detecting and tracking living beings during floods, leveraging their proven capabilities.

Real-time Data Analysis: Several reviewed papers emphasize the importance of real-time processing. This influences the project's objective to develop a system capable of rapidly analyzing visual data from flood-affected areas, enabling timely response actions.

Integration of AI in Disaster Response: The literature review showcases the potential of AI, specifically computer vision, in disaster response. This influences the project's objective to augment traditional methods with artificial intelligence, enhancing the efficiency of search and rescue operations during floods.

Enhanced Human Detection Techniques: Papers addressing drowning detection and multiple human objects tracking highlight the need for accurate human detection. This influences the project's objective to enhance the system's ability to identify specific human body shapes, movements, and gestures during floods

Chapter 3

Feasibility Study and Requirements Analysis

3.1 Feasibility

3.1.1 Technical Feasibility

Software Feasibility

Availability of Required Software : Computer vision libraries (e.g., OpenCV, TensorFlow, PyTorch) are readily available and compatible with the chosen programming language, Python.
Algorithm Implementation : YOLO and ResNet algorithms for human detection can be implemented within the project's technical constraints.
Integration Compatibility: Seamless integration of software components for data processing, machine learning, and real-time communication can be performed.

Data Feasibility

Data Availability : Relevant training data for YOLO and ResNet models, specifically focusing on diverse flood scenarios are available.
Data Quality : Optimal quality of existing datasets to ensure accurate model training and validation has been ensured.

Hardware Feasibility

Computational Resources : Computational requirements for real-time processing and model inference have been evaluated. Hardware resources, such as GPUs, for efficient execution are available.
Camera Hardware Compatibility : Successfully assessed compatibility with various types of cameras (CCTV, drones) to capture visual data during floods.

3.1.2 Economic Feasibility

Cost-Benefit Analysis : Conducted a thorough cost-benefit analysis, considering expenses related to hardware, software licenses, development resources, and potential savings in disaster response efficiency.
Return on Investment (ROI) : Assessed the potential ROI, taking into account the project's contribution to minimizing flood-related damages and improving emergency response.

3.1.3 Time Feasibility

Project Duration : The Computer Vision Guided Life Detection during Floods is estimated to have a total duration of 25 weeks, including all phases from initiation to deployment.
Working Time Allocation : Considering a standard 5-day workweek and 8 hours per workday, the estimated working time for the project is approximately 1000 hours.

3.1.4 Legal Feasibility

Data Privacy and Compliance : Ensured compliance with data privacy regulations, especially when dealing with visual data that may involve identifiable individuals.

Intellectual Property : Confirmed that the project does not infringe on existing patents or intellectual property rights.

3.1.5 Operational Feasibility

User Training : Evaluated the ease of use for emergency responders and ensure that necessary training can be provided.
Integration with Existing Systems : Assessed compatibility and feasibility of integrating the proposed system with existing emergency response frameworks.
Scalability : Ensured the system's scalability to handle varying scales of flood incidents and support potential future expansions.

3.2 Project Requirements

3.2.1 Implementation Requirements

The implementation of the Computer Vision Guided Life Detection System necessitates several minimum requirements to ensure the successful development and functionality of the proposed solution. These requirements include:

- Programming Languages: The implementation will utilize programming languages suitable for web development and machine learning tasks. This includes languages such as Python for machine learning algorithms, HTML, CSS, and JavaScript for the user interface.

- Web Framework: A web framework will be employed for the development of the user interface and system backend. Flask is specified as the chosen web framework to facilitate data processing, user management, and overall system communication.

- Machine Learning Framework: The implementation requires the use of a machine learning framework for developing and training object detection models.

PyTorch is identified as the chosen framework for building and training these models.

- Data Storage Solutions: The system interfaces with data storage solutions for the storage and retrieval of images and video streams. This may involve databases or cloud storage services, depending on the project's infrastructure.

- Communication Interfaces: Communication interfaces such as Wi-Fi, cellular networks, or other internet connectivity methods will be utilized to transmit alerts and notifications to users and relevant authorities.

- Development Tools: Access to development tools and libraries for machine learning model training, web development, and data annotation services is assumed. This includes tools such as data annotation platforms and machine learning libraries like PyTorch.

- Security Libraries: Software security libraries will be incorporated to ensure the secure transmission of data and implement user authentication mechanisms.

- Web Browser: Users must have access to a web browser for system access and interaction. The web-based nature of the application necessitates compatibility with standard web browsers.

- Privacy Compliance:The system must adhere to privacy regulations when handling sensitive data. The implementation should include measures to ensure the privacy and confidentiality of user information.

- Continuous Integration: The implementation should support continuous integration practices to streamline development processes, testing, and deployment.

3.2.2 Deployment Requirements

The deployment of the Computer Vision Guided Life Detection System involves considerations for ensuring the system's effective operation and accessibility. The deployment requirements include:

- Hardware Requirements: The system should be deployable on hardware configurations that support the processing demands of real-time object detection and machine learning algorithms. This may include servers, cameras, and drones.

- Operating System Compatibility: The system should be designed to work across different operating systems to enhance its portability. Compatibility with major operating systems such as Windows, Linux, and macOS is desirable.

- Web Browser Compatibility: Users should be able to access and interact with the system through standard web browsers. The system should be compatible with popular browsers such as Google Chrome, Mozilla Firefox, and Safari.

- Network Connectivity: The deployment environment should have reliable network connectivity to facilitate real-time communication, data transmission, and the delivery of alerts and notifications.

- Scalability Considerations: The system should be scalable to accommodate potential increases in load and additional functionalities in the future. This includes the ability to handle a growing number of users and data.

- Security Measures: Deployment should include security measures such as encryption for data transmission and robust user authentication mechanisms to protect against unauthorized access.

- Documentation: Comprehensive documentation should be provided to guide administrators and users through the deployment process. This documentation should cover installation, configuration, and maintenance procedures.

- Training Data Availability: The deployment assumes the availability of data annotation services or tools for training the object detection model. Adequate training data is essential for the accuracy and effectiveness of the system.

- Compliance with Regulations: The deployed system must comply with relevant regulations and standards, especially those related to privacy and data handling. Any legal or regulatory requirements must be considered during deployment.

- User Training: Adequate training and support should be provided to users, especially emergency responders and system administrators. This ensures that users are proficient in utilizing the system effectively.

Chapter 4

Project Design

4.1 System Overview

The system is composed of several interconnected modules, each serving a specific role in the process of life detection. The Video Input Module accepts video input from various sources and standardizes these feeds for further processing. The Preprocessing Module enhances the video quality, performs object tracking, and augments data to prepare it for analysis. The Human Detection Module utilizes ResNet for feature extraction and YOLOv8 for human detection, providing accurate and real-time results. The Post-processing Module generates alerts based on detection results and sends them to relevant authorities. Additionally, an Ensemble Learning module combines predictions from multiple models to enhance overall detection accuracy.

The seamless collaboration between these modules ensures a robust and efficient system capable of accurately detecting human presence in real-time, particularly in flood scenarios. This project significantly contributes to the field of computer vision, with potential applications in disaster management and public safety.

4.2 System Architecture

4.2.1 Architectural Design

The high-level architecture of the system is illustrated in the figure. It consists of the following key components:

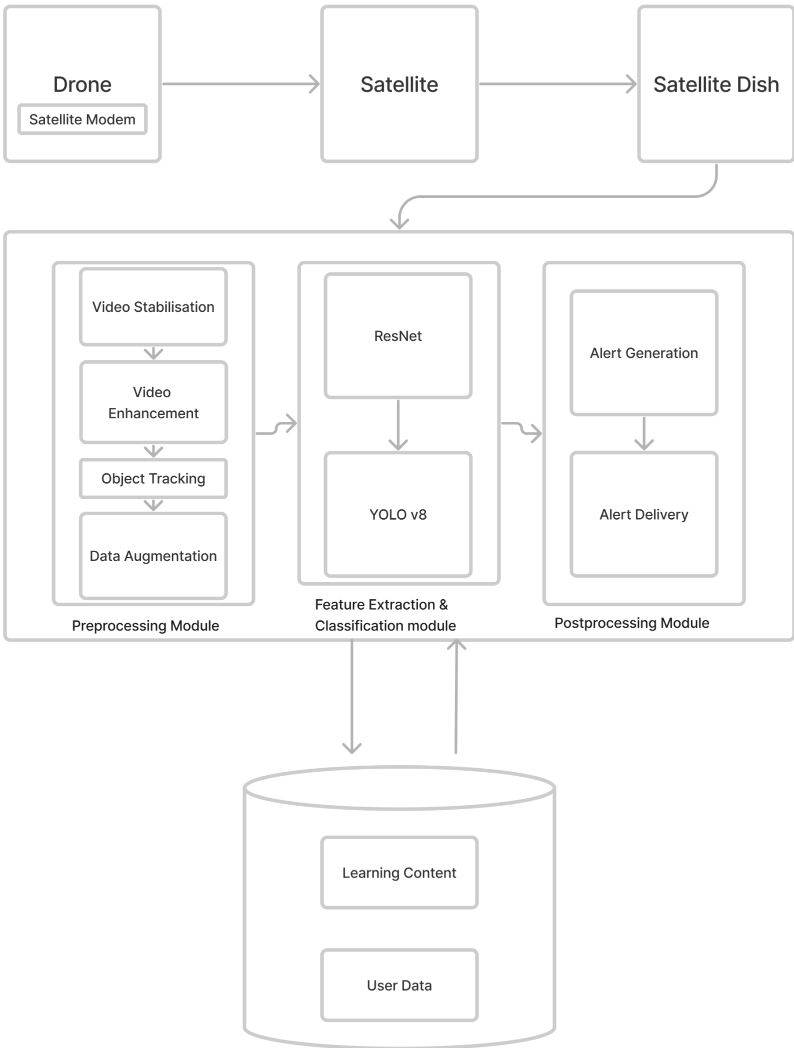

Figure 4.1: Architecture of life detection system.

Video Input Module: This module is responsible for accepting video input from various sources such as CCTV cameras, drones, or other recording devices.

It standardizes the video feeds for further processing and extracts metadata like timestamp and location.

Preprocessing Module: Upon receiving standardized video feeds, this module performs essential preprocessing tasks. These tasks include video stabilization to reduce camera shake, enhancement of video quality, object tracking, and data augmentation. The preprocessed video is then forwarded to the next stage.

Human Detection Module: This module is the core of the system, employing a ResNet model for feature extraction and a YOLOv8 model for human detection. It processes the preprocessed video, providing detection results, including the coordinates of bounding boxes around detected humans and confidence scores.

Post-processing Module: Once humans are detected, this module generates alerts containing crucial information such as the location of the detected human and the timestamp of the detection. The alerts are sent to relevant authorities through a reliable communication protocol.

Ensemble Learning Module: To enhance detection accuracy, this module combines predictions from multiple models. The ensemble learning approach ensures robust performance and reduces false positives.

4.2.2 Decomposition Description

- The **Video Input Module** functions as follows:

  Acceptance of Video Input: This component receives video input from diverse sources.
  Standardization: It standardizes the video feeds to ensure uniformity in subsequent processing stages.
  Metadata Extraction: Extracts metadata, including timestamp and location

data, from the video input.

Interactions: Sends standardized video feeds to the Preprocessing Module.

- The **Pre-processing Module** carries out the following tasks:

  Video Stabilization: Reduces the effect of camera shake to ensure stabilized video.

  Enhancement: Adjusts parameters such as brightness, contrast, and color balance to enhance video quality.

  Object Tracking: Utilizes object tracking algorithms to track and follow objects in the video.

  Data Augmentation: Applies transformations like flipping, rotating, and cropping to augment data diversity.

  Interactions: Receives video feeds from the Video Input Module and forwards preprocessed video to the Human Detection Module.

- The **Human Detection Module** is the core of the system:

  Feature Extraction: Uses a ResNet model to extract features from the pre-processed video.

  Human Detection: Applies a YOLOv8 model to analyze features and detect humans. Provides coordinates of bounding boxes and confidence scores.

  Filtering: Filters detections based on confidence threshold to reduce false positives.

  Interactions: Receives preprocessed video from the Preprocessing Module. Sends detection results to the Post-processing Module and model predictions to the Ensemble Learning Module.

- The **Post-processing Module** handles alert generation:

  Alert Generation: Generates alerts when a human is detected.

Information Inclusion: Includes information such as location, timestamp, and confidence score in the alert.

Communication: Sends alerts to relevant authorities using a reliable communication protocol.

Interactions: Receives detection results from the Human Detection Module.

- The **Ensemble Learning Module** improves detection accuracy:

  Combination of Predictions: Combines predictions from multiple models to enhance overall accuracy.

  Reduction of False Positives: Mitigates false positives through a collaborative prediction approach.

  Interactions: Receives model predictions from the Human Detection Module.

4.3 Design Rationale

The project is designed to detect human life in real-time during flood situations, potentially saving many lives. The system leverages advanced computer vision techniques, integrating ResNet for feature extraction and YOLOv8 for object detection. It adopts a modular design with separate modules for video input, preprocessing, feature extraction, human detection, post-processing, and alert generation, enhancing scalability and extensibility. Upon detecting a human, the system generates an alert and sends it to the relevant authorities. An ensemble learning module is also included to improve detection accuracy by combining the predictions of multiple models.

4.4 Data Design

4.4.1 Data Description

Video Data

This is the primary input for your system. It can come from various sources such as CCTV cameras, drones, or other video recording devices. The video data would typically include timestamp information, location data (if available), and the video frames themselves.

Preprocessed Video Data

This data is generated by the Preprocessing Module. It includes the stabilized and enhanced video frames, object tracking information, and any other data generated during preprocessing.

Feature Data

This data is generated by the Feature Extraction Module. It includes the features extracted from the preprocessed video using the ResNet model.

Detection Data

This data is generated by the Human Detection Module. It includes the detection results from the YOLOv8 model, such as the coordinates of the bounding boxes around detected humans and the confidence scores of the detections.

Alert Data

This data is generated by the Post-processing Module. It includes the alerts generated when a human is detected, along with any relevant information such as the location of the detected human and the timestamp of the detection.

4.4.2 Data Flow Diagram

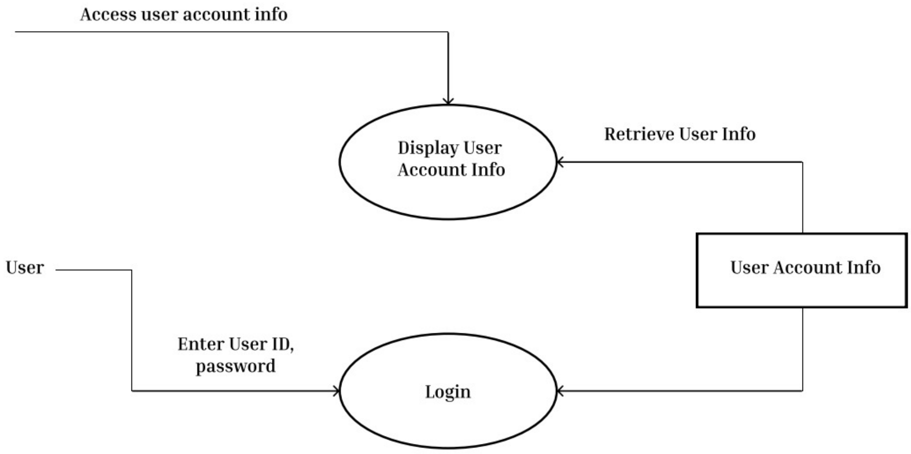

Figure 4.2: DFD for Login Subsystem

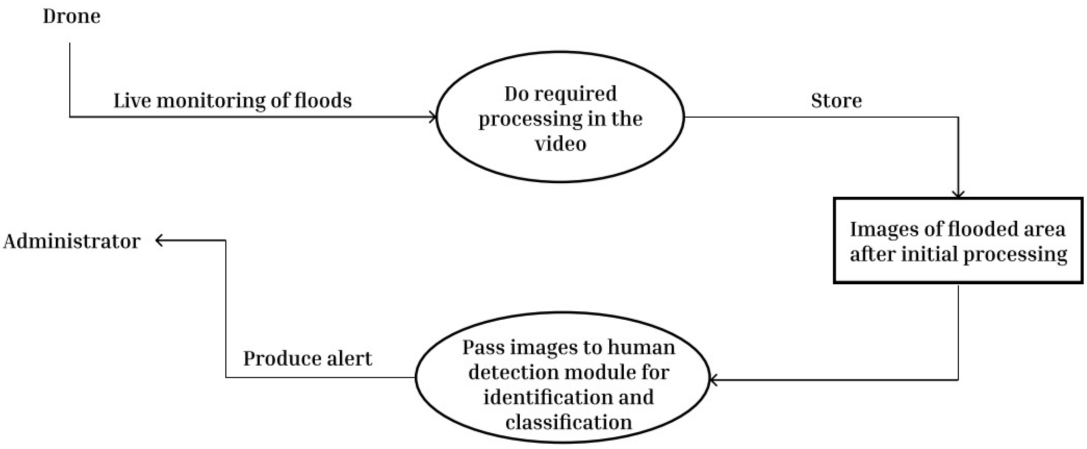

Figure 4.3: DFD for human detection and alert

4.5 User Interface Description

4.5.1 Administrator Registration page

The Administrator Registration page is the initial step for system setup. It features a straightforward design with input fields for the administrator's essential information, including username, password, and contact details. The form incorporates secure password requirements and validation checks to ensure the creation of robust administrator accounts. Clear labels and tooltips guide the user through the registration process, fostering an intuitive and user-friendly experience. The interface prioritizes data security and accuracy, setting the foundation for secure access to the system.

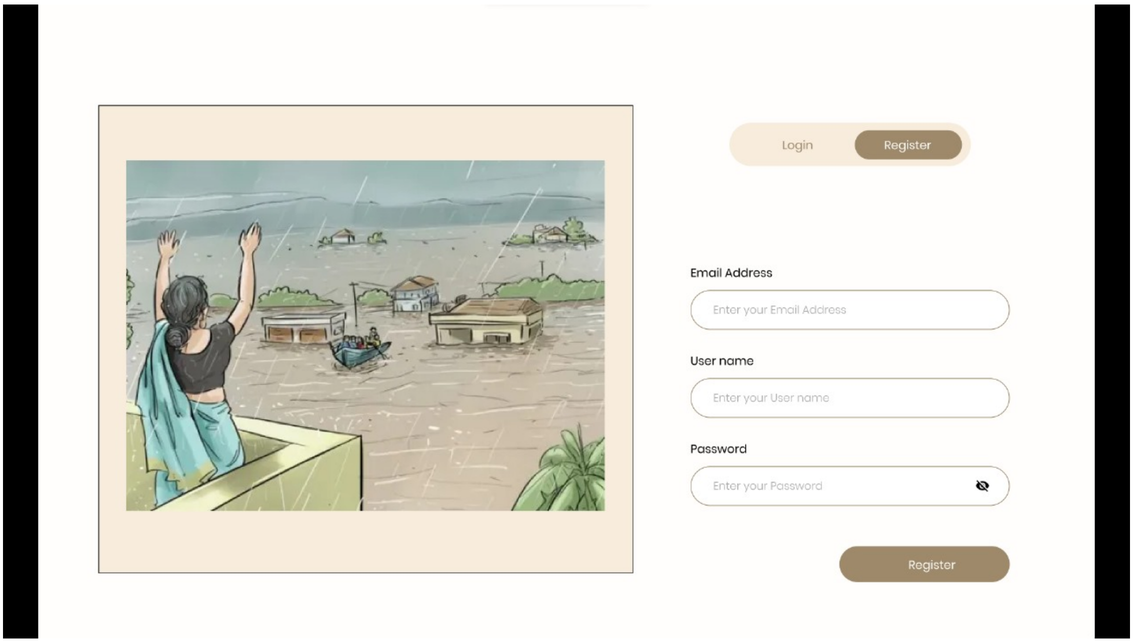

Figure 4.4: Administrator Registration page

4.5.2 Administrator Login page

The Administrator Login page provides a secure gateway for authorized access to the system. Users are prompted to enter their registered username and password, and a streamlined design ensures a quick and efficient login process. The interface includes error messaging to assist users in case of login failures and features password recovery options for added convenience. Upon successful login, administrators are directed to the system's home page, where real-time alerts and critical information are displayed, ensuring a seamless transition from authentication to system monitoring.

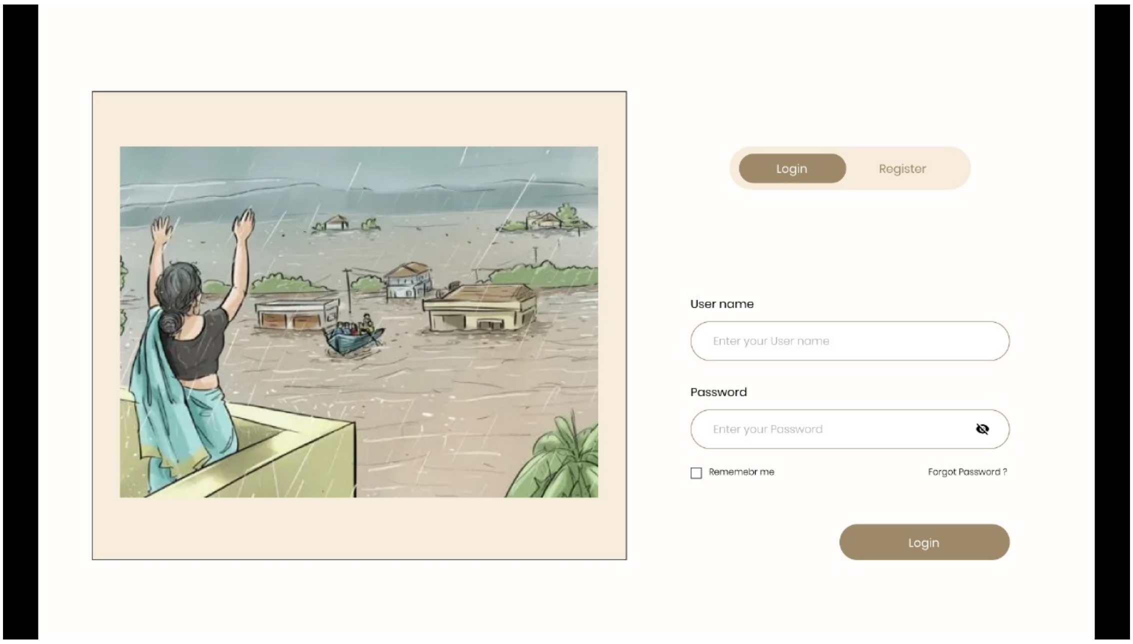

Figure 4.5: Administrator Login page

4.5.3 Home Page with Human Recognition Alert

The Home Page serves as the central dashboard for administrators, offering a real-time overview of the system's status. The interface prominently displays alerts when a human presence is recognized by UAV surveillance. A clean and organized layout ensures administrators can quickly grasp the situation, with timestamped alerts and location details. The design emphasizes visual clarity and responsiveness, enabling administrators to react promptly to detected human activity. Navigation options are intuitively placed, allowing administrators to access detailed information and take necessary actions with ease.

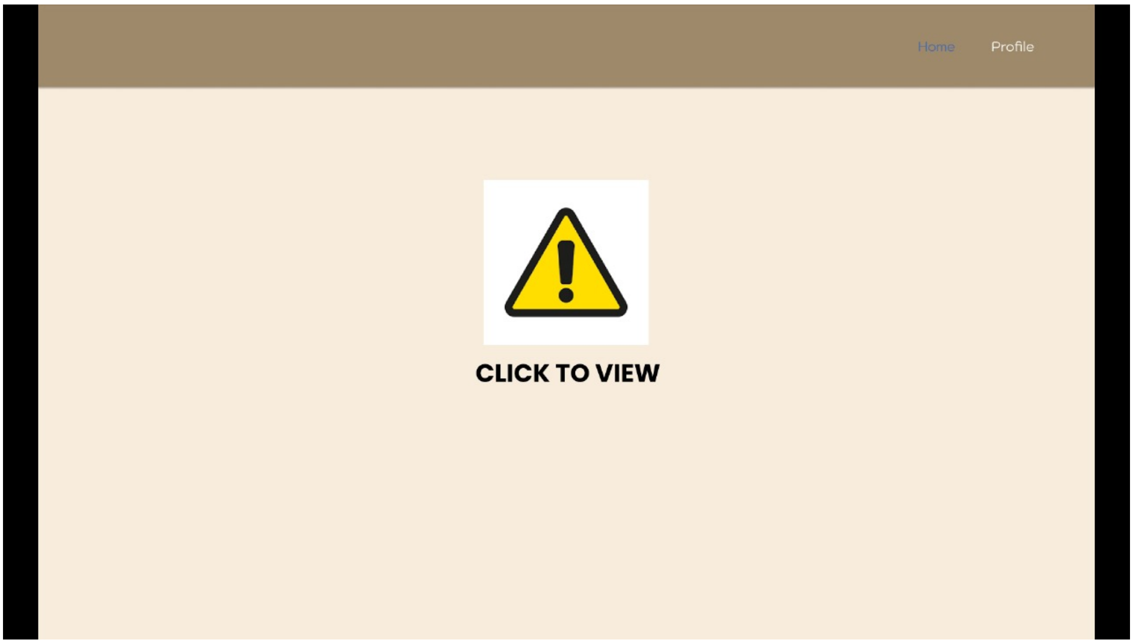

Figure 4.6: Home Page with Human Recognition Alert

4.5.4 Human Detection Details Page

The Human Detection Details page provides administrators with a comprehensive view of the detected human, featuring the captured image and precise location information. The interface includes a high-resolution image display alongside metadata such as timestamp, UAV coordinates, and any additional relevant data. Clear labels and annotations enhance the understanding of the presented information. Administrators can utilize interactive elements to zoom in on the image or access historical data. This page ensures that administrators have the necessary tools to make informed decisions and coordinate emergency response efforts effectively.

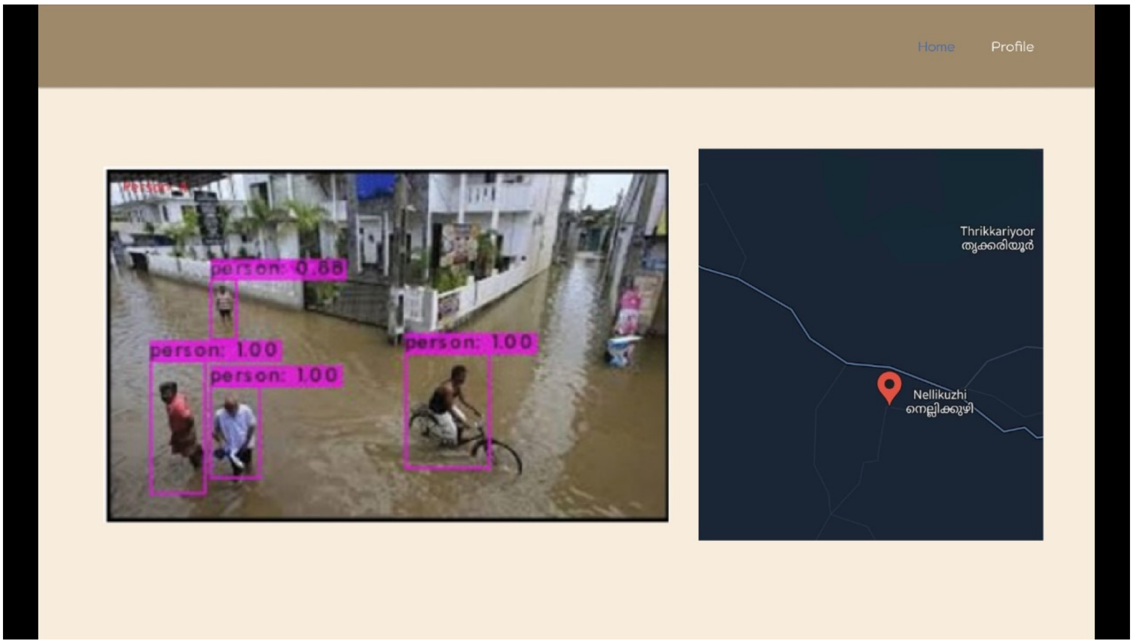

Figure 4.7: Human Detection Details Page

Chapter 5

Conclusions and Future Scope

5.1 Conclusion

The development and analysis of the Computer Vision Guided Life Detection System have led to several key conclusions:

5.1.1 System Performance and Effectiveness

The system has demonstrated robust performance in real-time human detection scenarios. Leveraging advanced computer vision techniques, the collaboration between modules, including feature extraction, object detection, and post-processing, has resulted in accurate and reliable identification of human presence.

5.1.2 Modularity and Scalability

The modular architecture of the system enhances development, testing, and maintenance processes. Each component operates independently, allowing for easy integration and facilitating scalability. This modularity not only supports the current requirements but also positions the system for future enhancements and adaptations.

5.1.3 Reliability in Challenging Environments

The system's design, incorporating video stabilization, data augmentation, and ensemble learning, contributes to its reliability in diverse environmental conditions. The ability to handle video input from various sources, coupled with efficient preprocessing and feature extraction, ensures the system's effectiveness in critical scenarios such as flood detection.

5.1.4 Contribution to Computer Vision

The project contributes to the field of computer vision by integrating state-of-the-art models like ResNet and YOLOv8. The utilization of these models for feature extraction and human detection showcases the potential of advanced vision techniques in real-world applications, particularly in disaster management and surveillance.

5.1.5 Collaboration with Authorities

The post-processing module's capability to generate alerts and communicate with relevant authorities adds a crucial layer to the system's utility. This feature ensures that human presence is promptly reported, facilitating timely responses in emergency situations.

5.2 Future Scope

While the current system fulfills the defined objectives, there exist several avenues for future exploration and improvement:

5.2.1 Enhanced Model Training

Continued research and development in model training, including the exploration of larger and more diverse datasets, can further improve the accuracy and generalization of the system. Fine-tuning existing models or incorporating newer models can contribute to even more reliable human detection.

5.2.2 Integration of Additional Sensors

Expanding the system to integrate data from additional sensors, such as thermal imaging or sound detection, can enhance the overall capability of life detection. This multi-sensor approach can provide a more comprehensive understanding of the environment, especially in challenging conditions.

5.2.3 Real-time Decision Support

Incorporating machine learning algorithms for real-time decision support can empower the system to not only detect human presence but also assist authorities in making informed decisions. This could involve risk assessment, evacuation planning, and resource allocation based on the detected information.

5.2.4 User Interface Enhancements

Improvements in the user interface, including visualization tools and interactive features, can enhance the user experience for administrators and emergency responders. Providing real-time dashboards, historical data analysis, and customizable alert settings can further streamline system usability.

5.2.5 Integration with Emergency Services

Exploring partnerships with emergency services and agencies can lead to the seamless integration of the life detection system into existing disaster response frameworks. This integration can facilitate faster and more coordinated responses during critical situations.

5.2.6 Edge Computing and Deployment

Investigating the feasibility of edge computing for on-device processing and deployment can be a valuable avenue. This approach can enhance the system's performance, especially in situations with limited connectivity or constrained resources.

References

[1] D. C. N. K. T. M. T. Do, M. H. Ha and Q. . H. D. Ba, "Human detection based yolo backbones-transformer in uavs," *2023 International Conference on System Science and Engineering (ICSSE)*, 2023.

[2] X. Y. T. He and M. Wangy, "An improved swimming pool drowning detection method based on yolov8," *2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC).*

[3] Y. Li and L. Liu, "Yolo-resnet: A new model for rebar detection," in *2021 4th International Conference on Robotics, Control and Automation Engineering (RCAE)*, 2021, pp. 128–132.

[4] A. Unnikrishnan, A. T. Roshni, P. R. Anusha, A. M. Vinny, and C. K. Anuraj, "Identification of drowning victims in freshwater bodies using drift prediction and image processing based on deep learning," in *2021 International Conference on Advances in Computing and Communications (ICACC)*, 2021, pp. 1–5.

[5] C. Kwan, D. Gribben, and T. Tran, "Multiple human objects tracking and classification directly in compressive measurement domain for long range infrared videos," in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2019, pp. 0469–0475.

[6] I. Chand, S. Kumar, J. Popli, and J. Kumar, "A stable deep learning model for recognizing human actions in noisy videos," in *2023 IEEE World Conference on Applied Intelligence and Computing (AIC)*, 2023, pp. 575–581.

[7] S. Bhatia, T. Chauhan, S. Gupta, S. Gambhir, and J. H. Panchal, "An approach to recognize human activities based on convlstm and lrcn," in *2023 6th International*

Conference on Information Systems and Computer Networks (ISCON), 2023, pp. 1–6.

[8] S. Güney and B. Erdaş, “A deep lstm approach for activity recognition,” in *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, 2019, pp. 294–297.

[9] David, G. Chiman, and T. Tran, Kwan, “Tracking and classification of multiple human objects directly in compressive measurement domain for low quality optical videos,” in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, 2019, pp. 0488–0494.

[10] S. Y. Nikouei, Y. Chen, S. Song, R. Xu, B.-Y. Choi, and T. R. Faughnan, “Real-time human detection as an edge service enabled by a lightweight cnn,” in *2018 IEEE International Conference on Edge Computing (EDGE)*, 2018, pp. 125–129.