# Abstract

With the increased digital transformation and the emergence of novel digital technologies, tremendous advancements have been seen in the field of computer vision and object detection. Still, the process of accurate object recognition and localization with a faster speed is a very challenging task. Human eyes can easily detect many objects in different videos or images regardless of the appearance of those objects but for computer systems and other digital devices, it is very difficult to recognize and differentiate different things or objects. However, the emergence of ML (machine learning) and DL (deep learning) methods optimizes these systems’ capability to detect different objects. However, these systems still pose poor accuracy during detection and also cannot detect small objects with higher accuracy and precision. This research work proposes an ensemble learning method and YOLO-based object detection model to overcome the existing limitations and enhance the accuracy of object detection tasks. The ensemble learning method is developed by combining ML and DL methods. Moreover, a Haar Cascade algorithm is used to improve the localization of the objects within the video streams and images. An NMS is used to recognize and remove the redundant bounding boxes and provide output as a single bounding box for each object present in the image. This proposed object detection model demonstrates a mean average precision of 98.678%, which is about 6.523% higher as compared to the normal YOLOv4 model.

Acknowledgment

I would first acknowledge the guidance, assistance, and support of my supervisor [name] throughout this research thesis. His guidance and support significantly helped me to better complete this research. I also want to express my gratitude to other teachers and staff members whose expertise contributed to completing this research. Finally, I want to provide heartfelt gratefulness to my family and friends whose love, support, and belief greatly helped to accomplish this milestone.

Dedication

I want to dedicate this research project to my family members, especially my elder brother, who supported and motivated me throughout this research work and assisted me in successfully completing this project.

# Introduction

The first chapter of this research focuses on introducing this research thesis, which is based on the development of an object detection system. As the world encounters a great emergence of digital transformation that revolutionizes many aspects. One field that is greatly influenced by this digital transformation is computer vision, which represents an important research field of AI (artificial intelligence) for deriving vital information or insights from videos or images (Simplilearn, 2023). The object detection process represents an important task in computer vision that includes real-time identification and localization of objects in images and videos while maintaining an appropriate level of accuracy. Object detection has enough potential to transform several industries by leveraging new products or services and automating various manual tasks. In this chapter, we will critically introduce this research project, including background information, problem statement, research questions, aim & objectives, description of the artifact, research rationale, and report structure.

## Background information

We can observe, distinguish, contextualize, and make sense of surroundings using a pair of biological cameras, namely eyes. However, they can’t record, store, and analyze the visual data but they can enable us to see different objects and enjoy the beauty around us. Let's imagine millions of eyes that can be placed in different places, such as buildings, streets, cards, satellites, etc., and using them we can get any desired information in the form of snaps to ease our lives in several ways. It can be made possible using computer vision (SuperAnnotate, 2021). Computer vision represents an interdisciplinary research field for enabling computers and systems to descend vital resourceful information from digital images and videos. It makes the computer systems or machines easily sense and understand the surrounding atmosphere by imitating human eyes and enabling the computer systems to perform several important functions using cameras, data, and algorithms instead of the retinas, optic nerves, and visual cortex.

Object detection represents a computer vision method to locate object instances in videos or images. Human eyes can easily detect and recognize presented objects in any image or video. The visual system of humans is very fast and can accurately perform complicated tasks, such as recognizing multiple objects and detecting obstacles easily (Kumari, 2023). With the increased evolution of faster GPUs, larger data amounts, and advanced algorithms, we can easily train computer systems for easily detecting and categorizing different objects present in any video or image with higher accuracy and precision. In this research, we will use machine learning and deep learning methods to develop a robust method that is highly scalable and effective to better deal with larger datasets and real-time applications. This proposed system will provide several features, such as accuracy, speed, scalability, robustness, and ease of use.

## Problem statement

This report mainly focuses on addressing the limitations faced by the existing systems used for object detection and improving their speed, accuracy, and performance for real-time applications. The currently used object detection methods face huge challenges in accurately recognizing and localizing objects, particularly in cluttered and complex scenarios (Contributor, 2015). These systems are not compatible with real-world applications due to their inability to attain the desired level of speed and accuracy. These systems cannot accurately identify and categorize the smaller objects due to difficulties in capturing the context and fine details of the smaller objects. Also, these systems are not effective in handling different environmental conditions, such as scale variations, occlusions, and diverse viewpoints. Thus, there is a crucial need to improve the accuracy & precision of object detection in reducing false negatives and positives. This research mainly focuses on implementing deep learning and machine learning methods for enhancing the accuracy of object detection in videos and images. It also intends to enhance the recognition of smaller objects, such as pedestrians and traffic signs within the videos or images using deep learning & machine learning methods with the YOLO algorithm to extract the restricted visual information (Jia et al., 2022). Moreover, this research will also provide critical insights into the decision-making processes to improve trust in the made predictions, especially for safety applications.

## Research questions

* How can we achieve accurate and precise object detection in real-time and compactly populated atmospheres?
* How can we enhance small object detection using the YOLO algorithm within the videos and images?
* How can the impact of class imbalance be mitigated using the YOLO and other DL algorithms without compromising accuracy?
* How can the developed YOLO-based object detection system enhance trust and the decision-making process, especially for safety applications?
* How can the proposed system be implemented with fewer computational resources while maintaining a reasonable speed and accuracy in edge-computing applications?
* How can the presented object detection system defend itself from confrontational attacks while ensuring security and reliability in sensitive applications?

## Aim and Objectives

### Aim

This project report mainly aims to develop an intelligent ensemble learning and YOLO-based object detection system to accurately and efficiently detect the diverse types of objects present in the videos and images.

### Objectives

* Addressing and reviewing the previously used object detection techniques and algorithms.
* Optimizing the capability of the object detection system to make it suitable for autonomous vehicles and video surveillance.
* Improves the detection ability of the object detection system for small objects to overcome the issue of limited visual information (Verschae & Ruiz-del-Solar, 2015).
* Developing a robust and easy-to-use object detection system that could be easily interpreted by humans.
* Assessing the performance measures of this presented system in terms of precision, recall, and mean Average Precision while comparing with other systems or methods.

## Description of this artifact

This artifact mainly focuses on addressing the issues or problems currently faced to accurately detect and classify the objects in the images and videos. It also focuses on exploring the research field of computer vision in object detection applications to enhance accuracy and performance. This system will be made available in the form of a website that can be used by users to accurately detect and classify objects from videos or images. This artifact will address the challenges related to its application in real-world scenarios, such as accurate object detection, enhanced detection of smaller objects, improved model trust & interpretability, and better handling of diverse environmental conditions. The core algorithm that is used for this presented system is the YOLO algorithm. Despite this algorithm, other ML and DL algorithms are also used. This introduced system takes input in the form of a video or image, and then the YOLO algorithm analyses this input and recognizes the presented objects in this input. This object detection system facilitates a real-time process of the video streams, interpretability & visualization tools, domain adaptation, and real-time deployment. This system can improve the accuracy of object detection and performance measures with increased security in different applications like video surveillance and autonomous vehicles.

## Research rationale

With the increased evolution of digital technologies, there is a need for a system to recognize and categorize diverse objects to enhance security on roads, buildings, and other places. The existing research studies proposed various object detection methods or systems but they are not very efficient and cannot provide accurate results (Patel, 2023). Thus this research mainly focuses on developing and introducing a novel YOLO-based object detection system to increase detection accuracy event for smaller objects. Moreover, the use of ML and DL methods helps in handling various issues, such as class imbalance, model deployment across different atmospheres, and enhanced transparency. This introduced system can facilitate real-time object detection across different real-world applications. Accurate and reliable detection of vehicles, individuals, and suspicious things can enhance security in homes, buildings, and public places. This research will also contribute to the advancement of the object detection process by implementing novel solutions, optimizing current methods, and promoting interpretability.

## Report structure

This project report is divided into six major chapters. The first chapter focuses on introducing this research project to develop a robust and efficient object detection system to accurately detect and categorize diverse types of objects present in the images and videos. Here I provide background information, problem statement, aim & objectives, research questions, description or artifact, research rationale, and report structure. The second chapter of this report will present a critical review of the previously published research papers or journal articles to understand other author’s works and used methods, theories, and concepts (Ehlers, 2017). The third chapter will discuss the research methodology, design, approach, and philosophy along with the other DL and ML methods, programming languages, and libraries. The fourth chapter will present the developed design of the introduced system along with its implementation. The fifth chapter will discuss the obtained research results, used testing methods, and a discussion of these results. Finally, the sixth chapter will conclude this research by summarizing key research findings and providing key directions to conduct future work.

Figure : Report structure

# Conclusion

In this part of the project report, this research thesis is critically introduced. It focuses on developing a YOLO-based real-time object detection system to accurately detect and classify multiple objects in complete scenarios. Here I provided background information, research questions, aim & objectives, problem statement, artifact description, research rationale, and report structure.

Chapter 2: Literature Review

# Introduction

This research part focuses on selecting and reviewing the existing research papers or journal articles that are relevant to this research project for gaining a better understanding of this research and attaining the most effective directions to conduct this research. Here the role of the YOLO algorithm will be explored along with the ML and DL methods in the accurate detection and localization of diverse objects in the images and videos. In this part, I will review the selected research papers based on the addresses research background, addressed research problem, proposed methods, concepts, & solutions, key contributions, and limitations. Further, the key research gaps and unfulfilled research requirements will also be discussed for future advancements and innovations.

# Literature review

## Object detection

According to Gallagher, (2023), object detection refers to a solution for identifying objects with their locations from an image or video. This object detection system returns the object coordinates from an image that is trained for recognition purposes. The main goal of this process is to identify and categorize different objects, such as animals, people, cars, or any other specific object in any given scenario. It includes two major steps, namely localization and classification. In the localization step, I accurately determine the object’s location in any video or image and tightly enfold the objects in the bounding boxes. After the location of the desired object, in the next step, it categorizes those objects into specific classes or categories and assigns a specific category or label to each object. There are various methods, such as computer vision, DL, and ML methods that can be used for attaining accurate object detection. The main applications of object detection include object tracking, autonomous vehicles, video surveillance, augmented reality, and robotics (Gallagher, 2023). The use of DL methods can significantly enhance the speed and accuracy of the object detection system while making them more effective and reliable for real-world scenarios.

## Role of Computer vision in object detection

State computer vision plays a vital role in the process of object detection by offering all the essential techniques and tools for better processing and analyzing visual data. The computer vision method includes the identification and localization of the objects in a video or images. It can facilitate a sophisticated object detection method for recognizing, locating, and classifying objects in real-world scenarios. It denotes an applied AI (artificial method) to provide an ability to the computer systems to efficiently perceive and interpret the visual world the same as humans (Patel, 2023). Object detection represents an advanced technology for training the machines to better recognize and understand the object’s characteristics and then apply it to various tasks, including the recognition of individual faces in crow and identification of traffic signs on busy roads. Some of the key roles that are performed by computer vision include feature extraction, object recognition, object localization, real-tie object identification, robustness, and integration ability with the other systems. Thus, suitable ML and DL methods are used, the computer vision can facilitate an efficient, accurate, reliable, and real-time categorization of objects in multiple domains or applications.

## Present and future directions in object recognition

According to Verschae & Ruiz-del-Solar, (2015), In the past few years, the world has faced a tremendous and successful expansion in the research field of computer vision. One of the reasons behind this increased emergence of the computer vision field is the advancements of ML methods. This research field comprises a process of object detection that can provide an ability for the robotic and computer systems to sense and capture visual information and categorize it. Object detection determines the scale and location of all the object instances, present in the considered image or video. Thus, the object detection system mainly focuses on identifying all of the object instances regardless of the location, scale, camera views, illumination changes, and partial obstructions. After the detection of object instances, these instances can be recognized, and tracked over an image sequence, and further information can be extracted from the object. This research discusses and compares several object detection methods, such as the coarse-to-fine & boosted classifier method, deformable part-enabled model, dictionary-based method, deep learning, and trainable image processing methods (Verschae & Ruiz-del-Solar, 2015). Object detection can be used in multiple applications, such as robotics, HCI (human-computer interaction), security, transportation, retrieval, and consumer electronics.

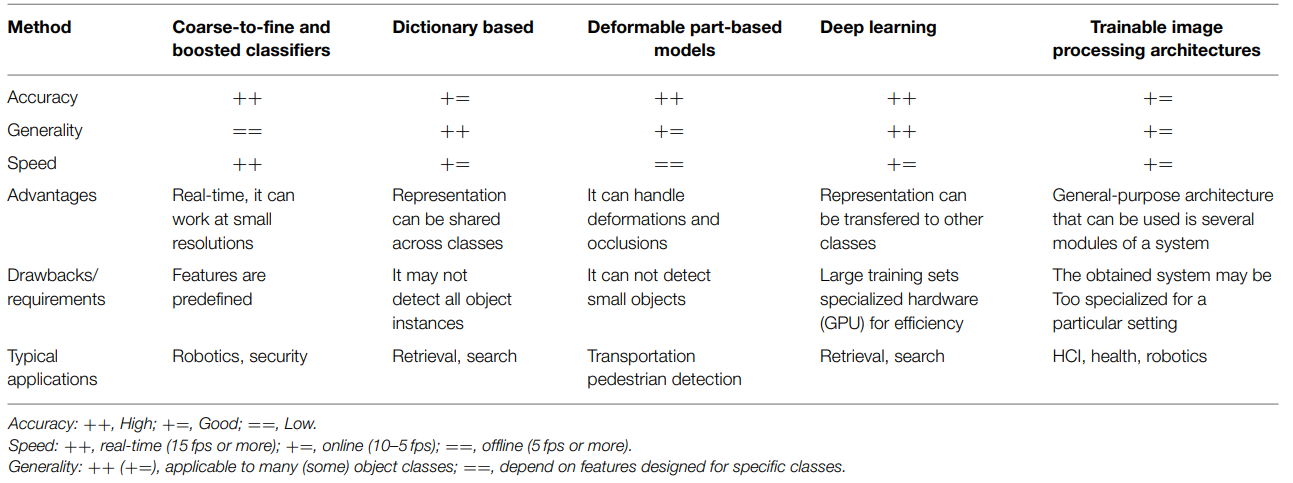


Figure : Comparison of the discussed object detection methods

## Object recognition and tracking methods

According to Panchal et al., (2015), the process of image processing represents a term that denotes the processing of a video or image frame that can be taken as an input and provide a specific set of relevant parameters with an image. This research addressed that the detection and tracking of the object become very important and it is widely used in the recognition of different objects in any image or video. Recently, human motion analysis become the most vital part of digital image processing to detect and sense human movement for detecting human motions from any background image or video. It is mainly performed to enhance an object’s existence in an image or video frame for detecting that particular object. This research proposed a background subtraction method shadow recognition theory for checking the presence of the objects in a video frame and identifying the objects. Then the identified objects could be categorized into multiple categories, such as vehicles, humans, animals, and so on (Panchal et al., 2015). However, the proposed method performed well it detected the objects with noise and also the obtained output was not accurate. Moreover, the objects behind any object were not detected.

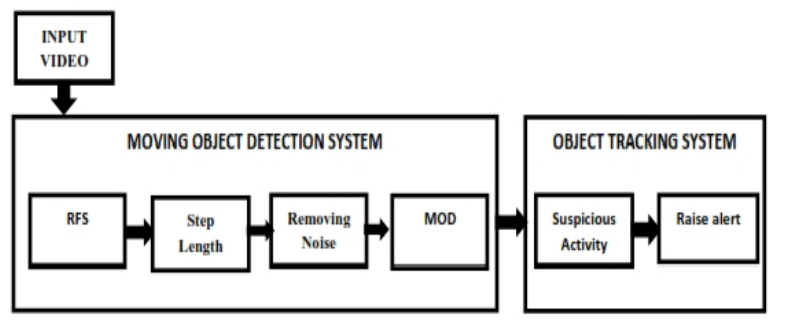


Figure : Block diagram of the object recognition and tracking

## Multi-sensor fusion system for moving object recognition and tracking

Cho et al., (2014) state that after the urban challenge of 2007 and the DARPA grand challenge of 2005, the researchers focused on developing exclusive opportunities for demonstrating the advances in autonomous vehicles. This study introduces a novel object detection system to detect and track moving objects. Autonomous or self-driving cars need to be able to consistently detect and track nearby objects in moving conditions before deployment in real-world driving scenarios. This research work introduces a novel recognition and tracking system for improving the previously used methods. The existing observation & motion models are revised for the active sensors, such as LIDARs and radars, and presented a new vision sensor. In this newly introduced system, the attached vision sensor module can detect vehicles, bicyclists, and pedestrians to generate more specific and corresponding vision targets. This presented system uses visually recognized information for enhancing the selection of the tracking model, classification movement, and data association of the existing system. Moreover, this model can realize effectual vision targets for improving the overall performance of the active sensor’s measurements. This proposed system faced a limitation in that it did not examine the obtained contextual information from the urban traffic atmospheres to further enhance the introduced system’s performance (Cho et al., 2014).

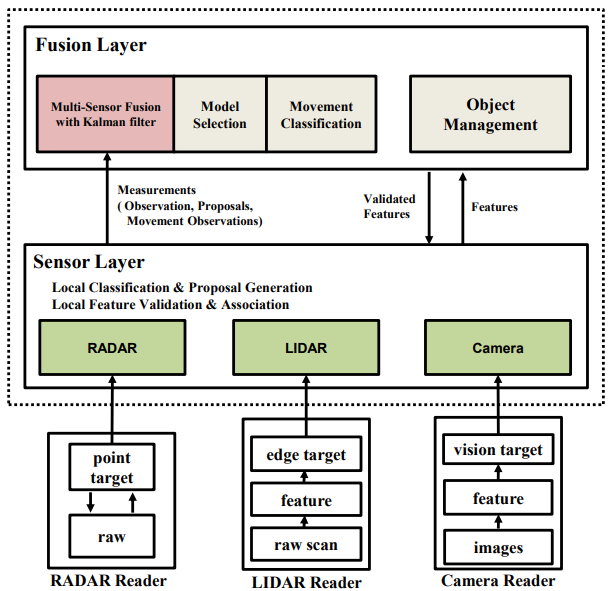


Figure : Block diagram of the introduced tracking system

## K-means clustering-based object detection to extract contextual information

Contextual information plays an important role in the computer vision and object detection process. The contextual information has been comprehensively deployed in the object recognition systems and can be utilized as the pre-descried spatial relationships. The spatial arrangements significantly manifest the heterogeneous statistical dissemination for diverse pairs of object classes that recommend class-stated prototypes of the spatial settings in a data-oriented manner. This study presents a new divergence K-means clustering method to automatically discover the prototypes of spatial contexts beyond the representations of the pre-described spatial relationships in the literature (Zhu et al., 2014). This research further constructs the vital features of spatial context following the learned prototypes with the help of a convenient localized soft assignment quantization approach. Further, to resolve the issue of complex spatial contextual features, this research presents a contextual occurrence-based K-L divergence and feature refinement technique for efficiently reducing the complexities in the proposed contextual model. The experimental outcomes conducted on the SUN 09 and PASCAL VOC datasets represent that the presented method can efficiently capture the most vital and meaningful spatial contextual prototypes along with the context feature related to other diverse class pais for boosting the performance of object recognition. This presented model cannot effectively integrate the local appearance and contextual information to realize a more sophisticated object recognition.

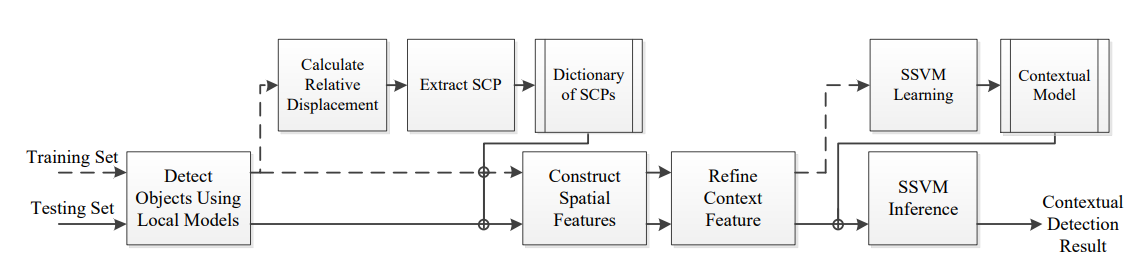


Figure : Overview of the introduced object detection system to extract contextual information

## Classification of contextual information obtained from LiDAR sensor

Niemeyer et al., (2014) argued that automatic and remote information extraction from the urban environment is one of the most challenging tasks due to complicated urban scenarios. This research work addresses the task of classification of contextual information from LiDAR cloud points. This research integrates the RF (random forest) classifier with the CRF (conditional random field) framework to realize a flexible strategy to obtain consistent classification outcomes. In this manner, it provides two benefits, the first benefit is the contextual consideration and another is an opportunity of using a larger feature set. This consideration of interactions helps to increase the accuracy by 2% by realizing the correctness and completeness of the discerned classes in the performed experiments. Further, this research compares the RF classifier with the other linear models for examining the relevancy of diverse features of the LiDAR sensor along with the interactions of neighborhood points. Then the building objects are recognized based on the categorized cloud point. The CRF class probabilities can be plugged into the Markov Random Field to include pairwise potentials. The experimental results demonstrate object detection with 100% completeness and 96% correctness (Niemeyer et al., 2014). This presented system poses a limitation in that it did not provide 3D classification and model the object interactions.

## Region-enabled CNN for an accurate object recognition and segmentation

According to Girshick et al., 2016), the recognition and localization of the objects in an image or video represent a most critical problem in the field of computer vision. However, with the increased use of lower-level image features, there is noteworthy progress has been made in it. The performance of object detection assessed on the official PASCAL VOC dataset has plateaued. The most effective methods are ensemble systems that can classically integrate several lower-level image features with the higher-level contexts. This research presents a novel convenient and scalable detection method for improving mAP(mean average precision) by 50% than the attained result using VOC2012, where a mAP of 62.4% is attained. The proposed methods combine two main ideas, the first is that the higher capacity CNNs (convolutional neural networks) are applied on the bottom-up province proposals for better localizing and segmenting the objects and the second is that the labeled training dataset is scarce and supervised to perform auxiliary tasks following a domain-centric fine-tuning process for significantly boosting performance (Girshick et al., 2016). The proposed method is also termed R-CNN (region-enabled convolutional neural network) because we combined the CNNs with the region proposals.

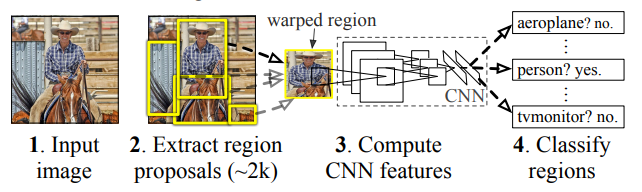


Figure : Overview of the proposed method

## ML-based intelligent vision system for autonomous object recognition

Ramík et al., (2013)addressed that the main problem in robotics is to design perceptual functions, so autonomous robots require perceptions for navigating in space and recognizing the environment and objects on which they are evolved. The previous recognition methods depend on the data labeled by humans which poses limitations in designing an autonomous machine vision system. This research work presents an advanced and intelligent machine vision system for better learning autonomous objects in real-time environments. This system heavily depends on salient object recognition. This system is inspired by the visual system’s early processing stages. Further, this research recommends a new and faster process to facilitate robust salient object recognition for various illumination conditions. Then it can be used for extracting salient objects to train the machine learning-enabled object recognition unit of this presented system. All the obtained results are provided on the MSRA Salient Object database to compare the performance of this presented method with other methods. Moreover, the presented system is implemented into a humanoid robot to increase its autonomy to better learn and interact with humans (Ramík et al., 2013). This research lacks a universally recognized algorithm that well-suits the proposed object detection method.

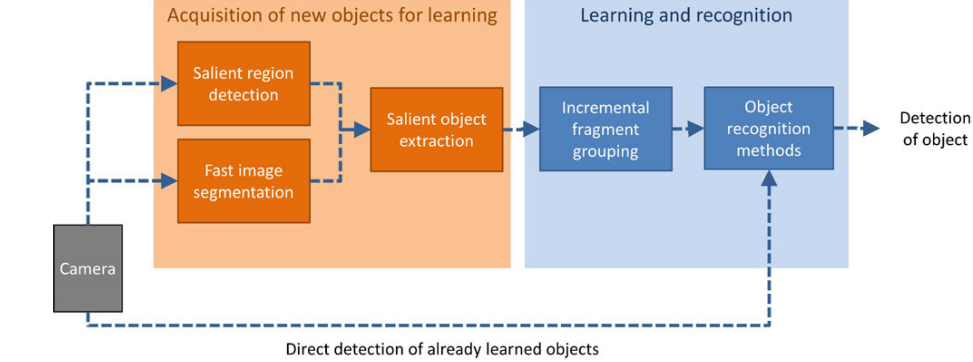


Figure : Block diagram of the presented system and its key units

## Pelee, a mobile-based real-time object detection system

Wang et al., (2019) address an increasing requirement of the CNNs (convolutional neural networks) on mobile phones with restricted memory resources and computing power motivates to development of an efficient design. In the past few years, several architectures have been present, such as MobileNetV2, MobileNet, and ShuffleNet. However, all of these presented models largely depend on the separable convolutions that lack an effective implementation in the DL frameworks. This study proposes a novel and efficient architecture, namely PeleeNet built on the traditional convolution. This presented PeleeNet system attains about highest accuracy on the ImageNet ILSVRC 2012 dataset and it is observed that it attains about 1.8 times faster speed as compared to the MobileNetV2 and MobileNet on the NVIDIA TX2. Also, PeleeNet is only 66% of MobileNet’s model size. Further, a real-time object recognition system is proposed by combining an SSD (single-shot multi-box detector) approach with the PeleeNet to optimize it for faster response. This presented system attain about 22.4% mAP on the MS COCO dataset and 76.4% mAP on the PASCAL VOC2007 dataset (Wang et al., 2019). The obtained results on the COCO dataset outperformed others with an 11.3 times small model size and 13.6 times low computational cost.

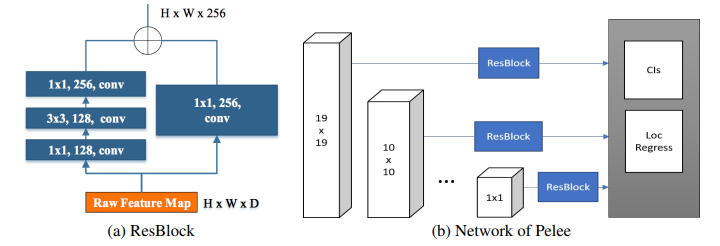


Figure : Architecture of presented Pelle system

## UnitBox, a modern object recognition system

Visual object recognition can be considered as the combination of two main tasks, such as object localization and visual detection. The deep CNNs (convolutional neural networks) witnessed a huge breakthrough in the detection of visual objects. This research utilized deep CNNs to better predict the object candidate’s bounding boxes and attained performance advantages over conventional regional proposal approaches (Yu et al., 2016). The currently used deep CNN techniques undertake the object bounds as the 4 independent variables that can be separately retreated. A generalized assumption conflicts with the properly received observations of the correlated variables and might result in less adequate localizations. Thus, this study presents a new IoU (intersection over union) function method for addressing this issue for the prediction of the bounding box, which reverts 4 bounds of the predicted box as the entire unit. This research introduced an intelligent object recognition system, namely UnitBox by taking advantage of deep CNNs and IoU loss function for performing an efficient and accurate localization to demonstrate robustness for the objects of varied scales and shapes with faster convergence. This research applied this presented system to the task of face recognition and attained an optimal performance among all of the other methods on the FDDB benchmark.

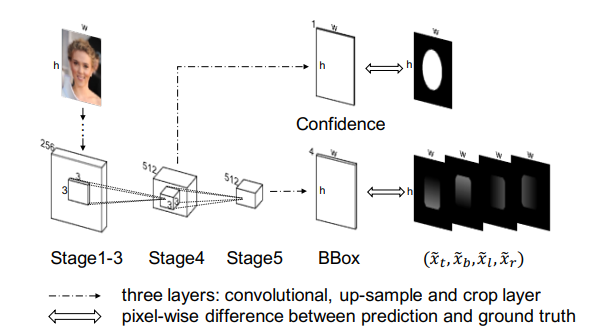


Figure : Architecture of the introduced UnitBox system

## T-CNN: Tubelets with the CNNs for object recognition from videos

The increased evolution the deep learning methods, such as deep CNN (convolutional neural network) and other object recognition frameworks significantly improve the performance measures of object recognition. In the past few years, the effectiveness and performance of these systems have been specifically optimized. Despite the evolution of robust deep neural networks like VGG and GoogleNet, modern object recognition frameworks, like R-CNN and its inheritors faster R-CNN and fast R-CNN, play an important role in enhancing the efficacy of conventional object detection systems. However, the existing studies only demonstrate its effectiveness on static images but they did not focus on videos (Kang et al., 2018). Also, they did not fully examine and utilize the contextual and temporal information of the videos. This research work proposes a novel DL (deep learning) framework for incorporating both contextual and temporal information obtained tubelets from the videos which extremely improves the baseline performance of the previously used object recognition systems from images. This system is called T-CNN (tubelets with CNNs) to recognize the objects from videos with higher accuracy. This research poses a limitation to not providing the details of the components of the proposed system to further enhance its accuracy and performance.

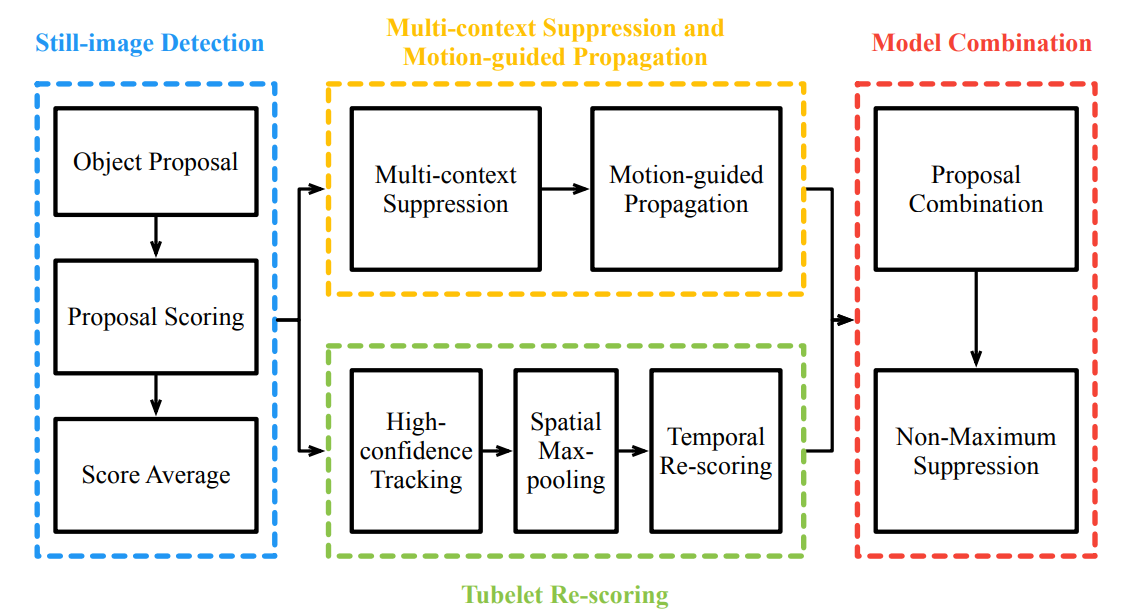


Figure : The proposed T-CNN framework

## DL methods for small object recognition

Object detection is the most important task of computer vision. A typical object recognition process includes two main steps, first is the localization of each targeted object and second is the categorization of those objects into diverse categories. Before the evolution of DL methods, object recognition techniques depended on the features and classifiers that were manually designed. There are significant advancements have been made in computer vision and object recognition with the deployment of deep CNNs. This study offers an inclusive review of the freshly developed DL methods for the purpose of small object recognition. This research first discusses the associated challenges and appropriate solutions for the recognition of small objects and suggests the best DL methods, such as adding contextual information, fusing feature maps, creating enough positives, and balancing the background-foreground examples (Liu et al., 2021). Further, the relevant methods with 4 main research areas, including facial recognition, generic object recognition, segmentation, and object recognition in aerial images will also be discussed. Moreover, this research compared the performance measures of the different DL methods in the recognition of smaller objects, such as SSD, YOLOv3, and Faster R-CNN following 3 larger datasets. The experimental outcomes demonstrated that Faster R-CNN provided the highest accuracy in the detection than other methods while YOLOv3 is second and DL methods pose the lowest accuracy. These methods pose limitations related to difficult size which prevents them from being deployed in mobile devices due to restricted computational resources.

## YOLO-based object recognition system in a traffic scenario

Tao et al., (2017) state computer vision represents a discipline to enables machines to look. It provides an ability for computer systems and other devices to behave as human eyes in the recognition of objects and better perform the task of image processing. The concept of computer vision and image processing become a modern research trend for extracting vital information from images like intelligent traffic surveillance systems. This research builds an efficient object recognition system for the images from traffic scenarios. It is an accurate, faster, and robust method. Conventional object detection systems first produce proposals and then extract features from those proposals. Then a specific classifier is executed on those proposals but its speed is slower and accuracy is also not sufficient. YOLO is an efficient and exceptional object recognition method, which is based on the DL methods and represents a CNN for localization and categorization purposes. All the connected layers of the YOLO system have been replaced with a typical pool layer to reproduce a novel network. Further, the loss function is augmented after the increment in the quantification of bounding coordinate errors. Moreover, a novel object recognition system, namely Optimized YOLO (YOLO) is generated which is 1.18 times more efficient than the general YOLO system and outperforms other region-oriented methods, such as R-CNN in terms of accuracy (Tao et al., 2017). Additionally, a combination of the R-FCN and OYOLO is added to the presented system. A histogram equalization method is used for pre-processing images at night.

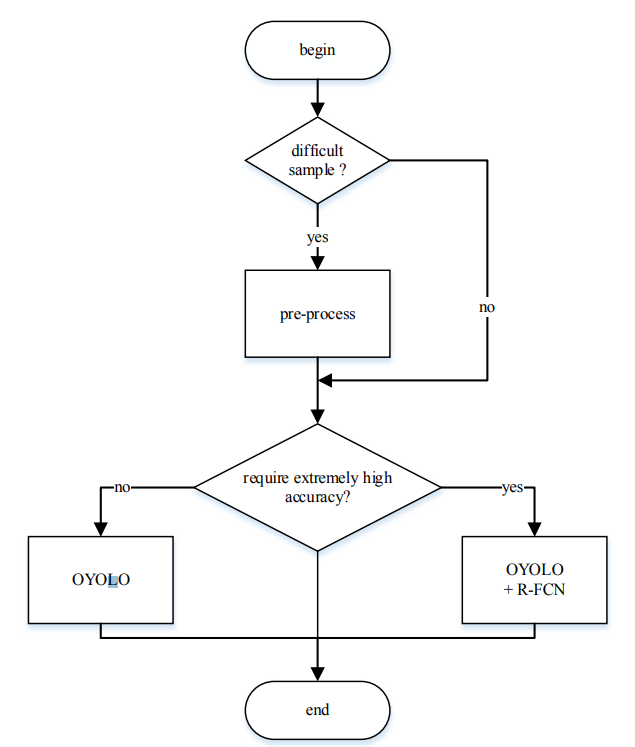


Figure : Entire process of the presented system

## Micro-YOLO for compressing object recognition model

Hu & Li, (2021) address that there is an accelerated growth in the research field of deep learning that largely endorsed the development and evolution of object recognition systems. It has comprehensive applications in robot vision, autonomous driving, facial recognition, and video surveillance. Conventional methods like YOLO and R-CNN create difficulties in deployment due to their complex design, tight budget, and partial computational resources. Thus, this research work presents a novel light-weight object recognition system based on CNN, namely Micro-YOLO to attain a noteworthy reduction in computational cost and the desired no. of parameters while maintaining an adequate level of recognition performance. This research also focuses on replacing the present convolutional layers with the depth separable convolution in the YOLOv3-tiny system, whereas the mobile reversed bottleneck convolution is replaced with the excitation & squeeze convolution for designing an advanced channel pruning method for reducing the no. of parameters and maximizing the performance of object detection, Thus, the presented Micro-YOLO system decreases the desired no. of parameters by 3.46 and MAC (multiply-accrue operation) by 2.55 then the YOLOv3-tiny system (Hu & Li, 2021).

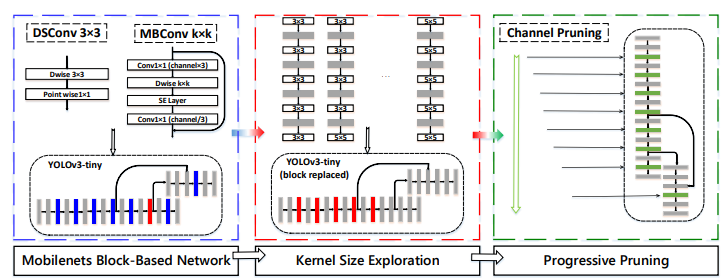


Figure : Architecture of the introduced Micro-YOLO

## Transfer learning-enabled YOLO system for detection of drain defects

According to Situ et al., (2023), the drainage system represents a key infrastructure that protects social assets and provides safety from polluted and excess water. This infrastructure system might suffer functional or structural damage that could directly impact the performance of this system and shorten its lifetime. Recently, the DL (deep learning) has demonstrated the most promising performance and effectiveness in the automatic recognition of defects in sewer systems. However, this methodology is data-centric and computationally exhaustive. The emergence of transfer learning significantly resolves the issues of data restrictions and also avoids the requirement of building models from scratch. This research study compared the performance measures of the presented transfer learning-enabled YOLO system with the other 4 mainstream object recognition techniques to recognize 5 main kinds of sewer. The experimental outcomes demonstrated that the introduced YOLO system outperforms the other object recognition techniques by increasing computation speed and improving the accuracy of detection. The other object recognition techniques can better recognize the disjoints but face difficulties in predicting the cracks and tree roots. Further, this research addresses the potential benefits of transfer learning and offers technical guidance for researchers and practitioners who need expertise in object recognition methods and accurate drain defect recognition systems (Situ et al., 2023).

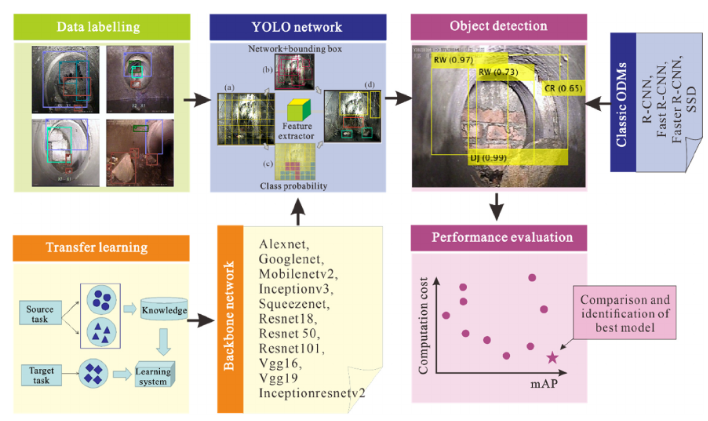


Figure : Workflow of the proposed methodology

## Haar Cascade classifier in the face recognition

According to Hashim & Mccullagh, (2023), the Haar Cascade classifier represents an efficient and popular method that utilizes ML methods in the detection of faces from digital videos and images. Face recognition represents a key component for several computer vision applications, such as image processing, security systems, and human-computer interaction. The Haar Cascade is considered the most efficient tool for face recognition because it can accurately recognize faces from videos and images even in complex atmospheres. In the process of face recognition, this algorithm utilizes a series of multiple classifiers trained on thousands of negative & positive images for identifying image regions that might comprise faces. This algorithm represents a multi-step procedure that includes the collection of training data, feature extraction, classifier training, and the development of cascade classifiers to detect faces from the test images. Then a post-processing of the obtained results is performed to remove the false negatives and false positives. This algorithm is extremely efficient and accurate in detecting faces from videos and images. Thus, Haar Cascade is a robust algorithm that is comprehensively used in the task of face recognition. However, there is a need to carefully assess its effectiveness in the context of each application and advanced methods should be used if required (Hashim & Mccullagh, 2023). However, it might face limitations in detecting faces under the condition of varying lighting conditions and when the faces are partially obstructed.

## Real-time detection of criminals using facial recognition & tracking methods

This research focuses on addressing the increasing rate of crimes and the limited presence of police by introducing an automatic and efficient facial detection system using the cascading classifier. This presented system deploys an automatic surveillance camera for better capturing real-time videos and smears facial detection methods for better identifying and detecting human faces. This introduced system uses s cascading classifier strategy for leveraging Haar features to recognize facial characteristics from the stream videos. This presented method can precisely identify and track individuals even in challenging environmental or lighting conditions (Somani et al., 2023). This system compares the processed face images with the available database of all known criminals to identify the criminal. If the system identifies a match, then it alerts the authority members with the location criminal to enable a quick intervention and response. This presented system poses more than 80% of the matching rate, which makes it extremely efficient and robust for the detection of criminals in public places. Further, this system can strengthen public safety and security by using modern face detection methods by recognizing and tracking individuals in real-time scenarios to ultimately enhance law enforcement efforts.

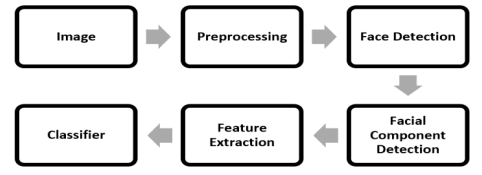


Figure : Workflow of the introduced system

# Conclusion

In this part of this report, I provided a detailed discussion of the conducted literature review of previously published research papers for addressing the other author’s works and their contribution to this considered research field. Here I critically reviewed about 18 research papers to create a robust foundation to further proceed with this research. Here I also addressed the research gaps in existing research works and key directions to further proceed with this research.

Chapter 3: Research Methodology

# Introduction

This research part is based on the discussion of the chosen research methodology along with the other tools, techniques, and methods to collect the most suitable research data and improve the accuracy and precision of object recognition. This section will realize an excellent roadmap for the entire research by outlining the taken steps to address the research questions and objectives. Here the selected research philosophy, research design, and research approach will be discussed along the justifying their selection. Further, the used programming language will be discussed along with the used libraries to conduct practical work and implement the introduced system.

# Research Methodology

Research methodology plays an important role in any research work or thesis. An effective and appropriate selection of research methodology can make any thesis successful. In this research, this research methodology will provide a well-structured framework to develop the presented YOLO-based object recognition systems while outlining the step-by-step process, such as data collection, data analysis, and ensuring an organized & system research philosophy and research approach.

## Mixed research methodology

In this project report, I have used a **mixed research methodology** to better conduct this research by exploring this research in both qualitative and quantitative aspects. The use of mixed research methodology helps me to use qualitative research methodology for providing an in-depth understanding of this research context along with addressing the user requirements and real-world issues, whereas the quantitative research methodology offers statistical evidence and vital quantitative insights (George, 2023). The quantitative research methodology includes a critical collection and analysis of the numerical data to offer the most vital statistical insights. In the development of a YOLO-based object recognition system, quantitative research methodology will enable an efficient performance assessment and optimization of this proposed system. The qualitative research methodology focuses on understanding the perspectives, behaviors, and experiences of individuals who are involved in object recognition system development.

### Justification

The use of mixed research methodology helps to realize and implement the vital research insights of both qualitative and quantitative research methodologies. The use of this research methodology also helps to combine both of these methods for strengthening the research design and research approach while offering the triangulation of the research findings and improving the research results credibility. The quantitative research method offers critical performance metrics, such as accuracy, precision, f1-score, and recall for evaluating the performance of the presented system, whereas the qualitative research method offers vital user perspectives and experiences. Thus, the use of mixed research methodology ultimately offers the most critical and valuable insights to enhance the usability, performance, and effectiveness of this presented object recognition system.

## Research design, research philosophy, and research approach

### Research design

In this research report, I used a **sequential explanatory research design** using which I first conducted quantitative research followed by qualitative research to provide vital research explanations and insights. This research design enables an in-depth understanding of this research based on the development of a YOLO-based object recognition system to enhance the system performance and user experiences (Ivankova et al., 2006). I first used the quantitative phase to focus on the evaluation of this introduced system and optimize this system. It includes data collection, model training, and performance analysis using key performance metrics, such as accuracy, precision, F1-score, recall, and mAP. It forms a solid base for the qualitative phase. In the qualitative phase, the relevant research works are reviewed to explore the other’s works, including used concepts & theories, limitations, and future enhancements in the presented YOLO-enabled object recognition system.

#### Justification

This research design enables an integration of the components of qualitative and quantitative research to realize complementary insights. The quantitative research phase offers performance assessment and optimization, whereas the qualitative research phase provides subject perspectives and experiences. The use of this research design helps to realize the potential of both quantitative and qualitative research for yielding performance metrics and numerical outcomes as well as explaining & interpreting research findings along with exploring the challenges, and improvements. Further, this research design optimizes the cogency and trustworthiness of the obtained research findings.

### Research Philosophy

I use **pragmatism research philosophy** for this project report based on the development of a YOLO-based object recognition system. This research philosophy emphasizes practical knowledge for addressing real-world issues. It mainly focuses on the utility and relevance of the research findings. In this research, this philosophy is used to integrate key insights of qualitative and quantitative research methods for offering an in-depth understanding of the development process and practical implications of the presented system. Further, it acknowledges the importance of subjective user experience and objective performance evaluation for bridging the gap between theoretical and practical knowledge. Moreover, it helped me adjust the adopted research approach following the evolving challenges and needs recognized throughout the entire development process.

#### Justification

I used this research philosophy because it offers an effective understanding of the associated user needs and practical challenges with this presented system to ensure the adoption and usefulness of this system. It motivates the researchers to consider the comprehensive ethical and societal impacts of the research work, which is essential for the responsible development and execution of the introduced object recognition system. Moreover, it can efficiently test and refine this presented solution in the context of the selected application domain.

### Research approach

For developing the YOLO-based object recognition system, I used a **concurrent embedded research approach** by considering quantitative research as the primary focus, where qualitative research on the supplementary component to gain in-depth understanding and insights (Bell et al., 2022). It integrated the qualitative and quantitative research findings to provide a better understanding of the development of the proposed YOLO-based object recognition system. It also strengthens the validity of the obtained research findings while ensuring the robustness of the drawn conclusions. It offers a detailed understanding of the strengths, weaknesses, and improvements concerning both subject user experiences and objective performance metrics.

#### Justification

I used this research approach because it helps to integrate and utilize both quantitative and qualitative research methods. It offers a number of potential benefits, first, it helps to better understand the user requirements and expectations by suggesting effective evaluation criteria and system design. Further, it helped in integrating the feedback of users from different backgrounds for identifying any potential biases and refining the proposed system to realize comprehensive applicability. Further, it will better explore the user privacy concerns, ethical considerations, and societal impacts of this presented system.

## Methods

### Machine learning methods

Machine learning methods play an important role in the development of the proposed object recognition system by deriving an ability to recognize and locate the objects within the considered images or video frames. The ML-based feature extraction methods help to better extract and learn the complicated features and insights from the images or videos. Then these methods help in categorizing the extracted features from the images and refining the made predictions to accurately localize the objects within the images. Further, it can be integrated with the YOLO and Haar Cascade methods to expand the presented system’s generalizability and resilience.

### YOLO algorithm

YOLO stands for You Only Look Once, which is a very effective and popular algorithm mainly used for object recognition. It divides an image or video into a grid system, where each grid perceives the presence of the objects within itself. In this project, we YOLOv7, which is the latest version of the YOLO algorithm launched by Ultralytics (Srivani, 2023). This YOLO algorithm is faster, simpler, and easier to use. It can detect objects from images and videos with higher accuracy and precision.

YOLO uses an end-to-end neural network for making the forecast of all the class probabilities and bounding boxes at once. It is different from all the existing methods used for object recognition and outperforms them in terms of accuracy and effectiveness. It performs predictions using a single and fully connected layer. This algorithm takes input in the form of an image and deploys a simple convolutional neural network for detecting the objects from images. First, an ImageNet is used to pre-train the convolutional layers by the plug-in into the fully connected layer. Further, this pre-trained model can be converted for performing the detection process while improving performance. Its fully connected layer can better predict both coordinates of the bounding box and class probabilities.

The YOLO algorithms first divide the input image into a grid of SxS dimensions and the object’s center falls into the grid cell, then this grid cell is responsible for detecting the object. Each grid cell can predict confidence scores and bounding boxes for those boxes (Kundu, 2023). These confidence scores imitate the model’s confidence that the object is present in the box. This YOLO algorithm can also predict many bounding boxes per cell of the grid. YOLO allocates responsible predictors to predict an object following the highest IoU of the predictions. It leads to the specialization among different bounding box predictors. An NMS (non-maximum suppression) technique is used by YOLO, which represents a post-processing stage for improving the efficiency and accuracy of object recognition. During the process of object recognition, it might be possible that bounding boxes may overlap and represent the same object. Here NMS helps in recognizing and removing the wrong or redundant bounding boxes and provides output as a single bounding box for each object present in the image.

### Haar Cascade algorithm

The Haar Cascade is one of the most simple and frequently used algorithms for detecting objects in images regardless of their location and scale in the image. It is not very complicated and can be run easily into real-time scenarios. It has a haar-cascade detector that can be trained to recognize multiple objects, such as fruits, bikes, cars, buildings, etc (Jaiswal, 2023). It utilizes a cascading window for computing the features in each window and categorizing to ensure whether it is an object or not. However, this algorithm can be easily implemented and needs minimal computing power but it is highly prone to false positives and needs parameter tuning. Thus, we used an OpenCV Haar Cascade algorithm to overcome these limitations.

### Product backlog approach

A product backlog approach is used to support the development process for the introduced YOLO-based object recognition system. This approach helped in first recognizing the relevant user stories and then prioritizing them to evaluate the desired efforts for each of these user stories. After this a product backlog is created, including the sprints for this proposed method. Further, the feedback is collected from the users. These all steps are repeated till the completion of the system development process. Then this approach is used to assess the proposed system’s effectiveness against the specified project goals and requirements (Boesch, 2024). Moreover, this approach lets the developers put more effort into training the developed system for gradually launching novel test data and making desired system requirements. Additionally, it offers a concise roadmap to develop this system following user requirements and getting early feedback from the users.

### Project management approach

In this research project, I used an Agile project management methodology, which is the most effective and suitable project management approach for this research. As the deep learning methods are unpredictable and very complicated, so it will provide an effective level of adaptability and flexibility. Using this project management approach, this entire research project can be divided into small and manageable parts called sprints (Laoyan, 2022). Further, it integrates user feedback at regular intervals to ensure that all the user’s preferences and needs are met. Moreover, the agile methodology also enables an appropriate adjustment of the research methodologies, goals, and system design following unforeseen obstacles or novel insights.

### Programming language and libraries

In this research project, I used Python programming language to develop the necessary code for developing and implementing the presented YOLO-based object recognition system. Currently, Python is the most popular and frequently used programming language, which is used in web development, app development, machine learning, desktop applications, and more fields. Python can be easily learned due to its easy and simple syntax, which makes it a great programming language for beginners. I first set a virtual atmosphere for isolating the project dependencies for these presented systems. Then I installed the required libraries, such as NumPy and OpenCV to better follow the running commands in the virtual atmosphere. A TensorFlow library is also used for loading and working with the YOLO algorithm based on the configurations and weights in Pytgon. The used libraries in this project are as follows.

### TensorFlow

TensorFlow represents a Python-friendly open-source library used for the development of neural networks and machine learning-based applications. This library helps to easily process the data collection, model training, make predictions, and refine the obtained results (Yegulalp, 2024). It enables the developers to create effective dataflow structures of graphs for better describing how much data is moving through that graph.

### OpenCV

OpenCV is the most popular and robust tool mainly used for image processing and computer vision tasks. It represents an open-source library that could be used for better-performing tasks, such as object recognition, landmark detection, object tracking, etc (Gupta, 2022).

### NumPy

NumPy represents The most fundamental package for the purpose of scientific computing. This library offers a multi-dimensional array object with multiple derived objects to make faster array operations, such as logical, mathematical, sorting, shape manipulation, basic statistical operations & linear algebra, and random simulations, and so on.

# Conclusion

This research critically discussed the use of mixed research methodology along with the sequential explanatory research design, pragmatism research philosophy, and concurrent embedded research approach. Then I discussed the YOLO algorithm, Haar Cadcad algorithm, and product backlog approach. Then the Agile project management methodology is discussed. Finally, the Python programming language along with the NumPy, TensorFlow, and OpenCV libraries is discussed.

Chapter 4: Design and Implementation

# Introduction

This part or chapter of this project report is based on the development and discussion of the design of the presented YOLO-based object recognition system to easily and accurately recognize objects from images or videos. Here a basic and detailed design of the presented system will be created and discussed along with the key components. Then the implementation of this presented system will be discussed along with the evidence of the performed practical work in the form of created codes. In this chapter, a detailed understanding will be provided of the presented systems throughout their design, development, and implementation. This research part will bridge the gap between the theory and practicality of this research.

# Design and Implementation

## Design

The design section is one of the most critical parts of this research because it assists in the development of a robust strategy for carrying out any research following active planning. A few components are comprised in the developed system design, these components could be converted into potential actions for creating a basic-level and advance-level architecture of the presented system (Sangeetha et al., 2023). This basic system architecture is depicted in Figure 14, which comprises four vital components, namely the user, YOLO model, object recognition system, and object detection results. The YOLO algorithm significantly improves its efficiency and accuracy in the recognition of different types of objects from any given image or video. The used libraries, such as NumPy, OpenCV, and TensorFlow create a robust benchmark to better implement this presented system and make this system more user-friendly.

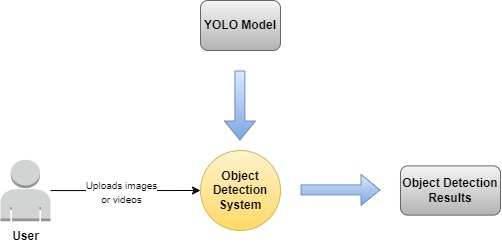


Figure : Basic architecture of the presented object recognition system

## Implementation

This research project is entirely based on the development of an efficient and accurate object recognition system with the help of ML and DL methods along with the YOLO and Haar Cascade algorithms to more efficiently and accurately detect the different objects from the given image or video as an input.

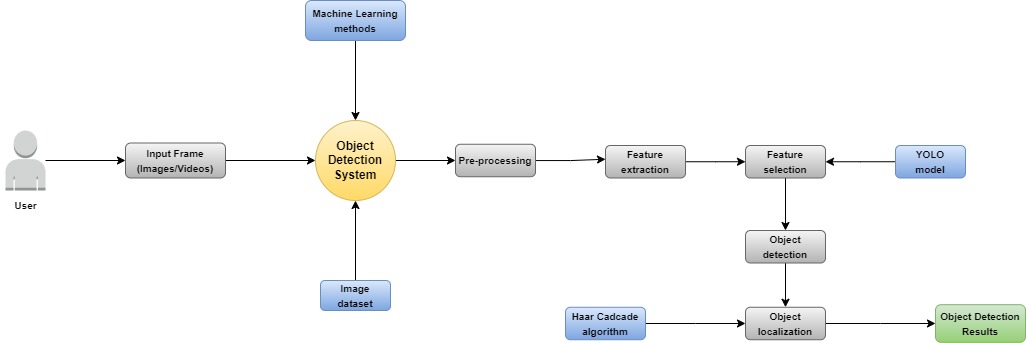


Figure : Advanced-level architecture of this presented object recognition system

Figure 16 depicts an advanced-level architecture of this presented object recognition system. According to this diagram, the proposed web-based object recognition system can be used by users. The users just need to upload any image or video to this system to the given input frame. This input frame is sent to the object detection system, where an image dataset is used from which the objects will be detected (Thangavel & Karuppannan, 2023). Here machine learning methods are used for pre-processing the given images or video as an input. Then vital features are extracted and selected from these images. Here YOLO algorithm is used to detect the objects as per the extracted features. Further, a Haar Cascade algorithm is used to accurately localize the detected object with higher classification accuracy to realize the final and most precise object detection results.

### Evidence of the performed practical work

The below-illustrated figure depicts the details of the executed commands and imported packages for running the scripts (Kristian et al., 2020). Here we imported FPS, VideoStream, NumPy, argparse, imutils, time, and cv2.

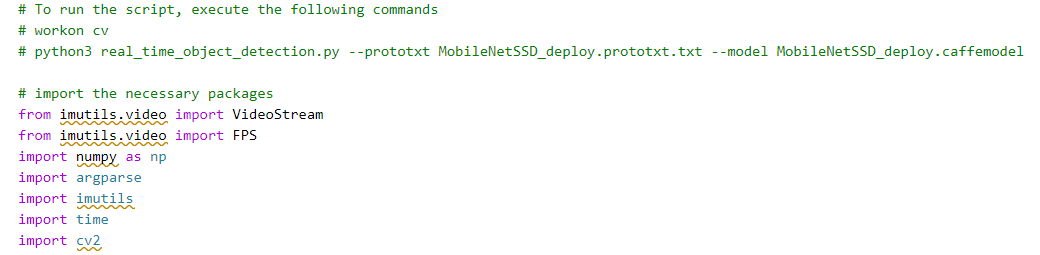


Figure : Running script and importing essential packages

The below illustrated figure represents the construction of an argument parser and parsing of the necessary arguments. Then a list of all the class labels from MobileNet SSD is trained for the purpose of object recognition and a set of bounding box colors is generated.

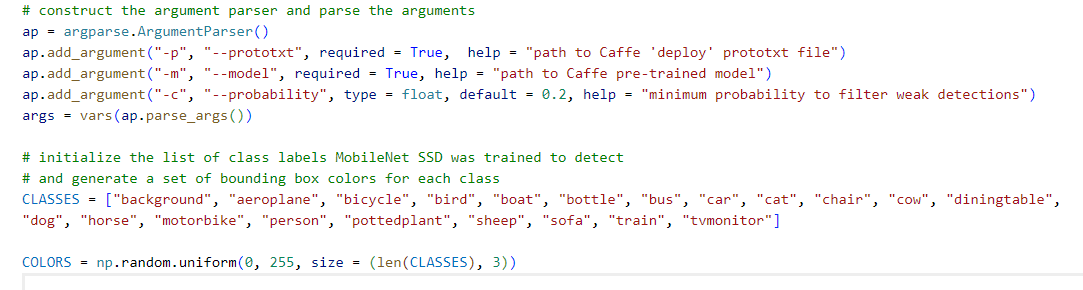


Figure : Constructing argument parser and initializing class label list

Figure 19 represents that a serialized model is loaded from the disk the video stream is initialized and the camera sensor is enabled for warming up. Then the FPS counter is initialized to further proceed with this process.

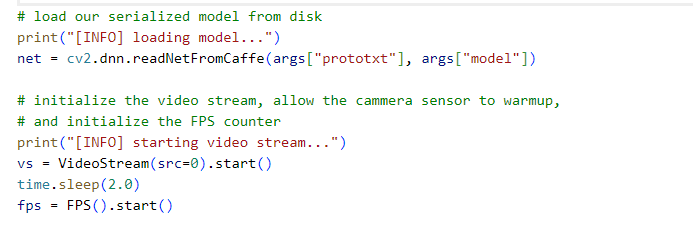


Figure : Loading serialized model and initializing video stream

Figure 20 depicts the present loops over the initialized video stream (Singh, 2019). The window of this video stream is resized to a maximum width of 500 pixels. Then the frame dimensions are grabbed and converted into a blob. Then this blob is passed via the network and performs the detection task. A loop over the detection outcomes is run to get the probabilities of made predictions.

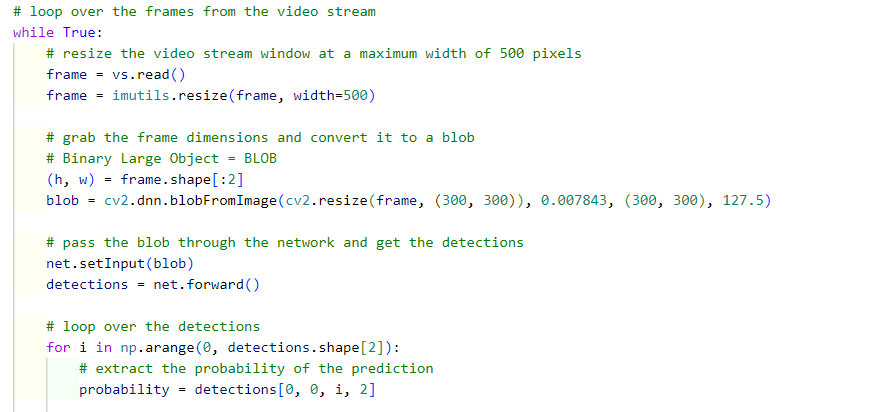


Figure : Presenting the loops over the initialized video stream

Figure 21 represents the filtering out process of the made weak object detection results by ensuring that only the detections that have a probability higher than the set min probability are included. Then the class label index is extracted from the detections and computed on the x & y coordinates of the object’s bounding box. Further, the predictions are drawn on the frame.

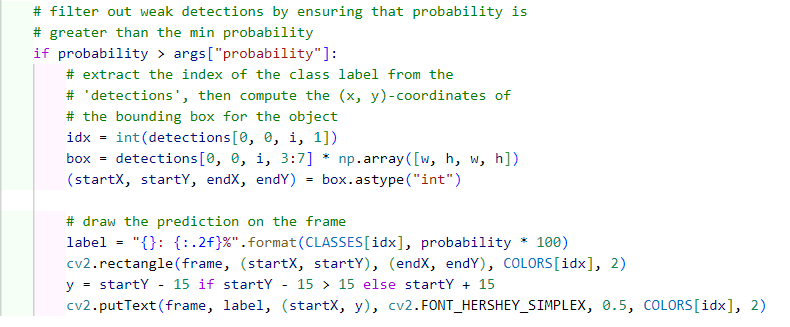


Figure : Filtering out the weak detection results

Then the output frame obtained from the detections is represented (Shamkuwar et al., 2022). When the ‘q’ key is pressed, then the loop breaks down and the FPS counted is updated.

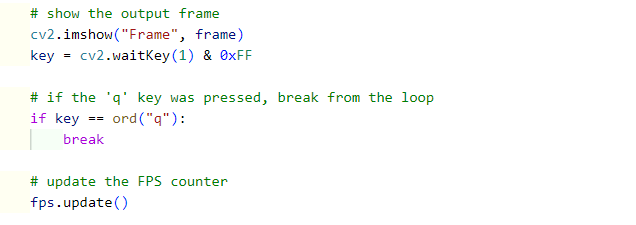


Figure : Showing output frame

After updating the FPS counted, the timer gets stopped and the obtained FPS information is displayed. Finally, a cleanup process is conducted to clean the redundant and wrong results.

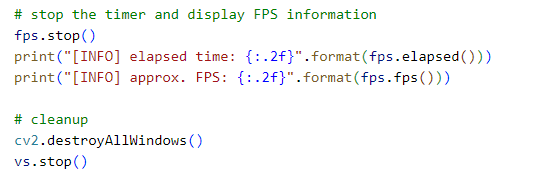


Figure : Stopping timer and displaying FPS information

# Conclusion

This research part is based on the discussion of the system design and implementation. Here I first discussed the basic level of the system architecture of the presented YOLO-based object recognition system with its key components. Then an advanced level of architecture is presented to demonstrate the process of its implementation. Then the evidence of the performed practical is presented in the form of created codes.

Chapter 5: Results, testing, and discussion

# Introduction

This chapter of this project report is based on the discussion of the obtained research results, used testing methods, and discussed on the attained research results. Here first the obtained experiment results will be presented with the experimental setup, performance metrics, and comparative analysis. Then I will discuss the used testing method to test the effectiveness of this proposed system and ensure that it performs as intended and meets all the user and business needs. Then I provided a critical discussion of these obtained research results along with deliberating the commercial and economic feasibility of this presented system.

# Results, testing, and discussion

## Results

The obtained research results are presented and the performance of the proposed system is validated with the help of a t-SNE visualization tool while addressing the extent to which the proposed system can extract rich insights and features from the images or videos. Then this visualization tool is used to observe the extracted features and map the data in higher to lower-dimensional order. The available categories in the taken data are not compatible with each other and they indicate that the differences between those categories are larger and easy to recognize. Also, these characteristics pose similarities, which might easily cause misconfiguration. However, this presented intelligence object recognition system is robust enough to extract vital characteristics but it is somehow inappropriate and should be further enhanced.

### Performance evaluation metrics

The performance evaluation metrics are important to reasonably evaluate the performance of the presented system from the perspectives of both precision and real-time (Yan et al., 2021). The performance metrics, such as accuracy, precision, f1-score, and recall are used to assess the effectiveness and performance of this algorithm.

Generally, there are 4 main relationship definitions for the negative and positive sample relationships. True Positive (TP) denotes that the positive samples are correctly recognized, True Negative (TN) denotes that the negative samples are correctly recognized, False Positive (FP) denotes that the positive samples are incorrectly recognized, False Negative (FN) denotes that the negative samples are incorrectly recognized.

A classical performance evaluation can be attained using a confusion matrix that could be created by arranging these 4 negative and positive samples in the form of a matrix.

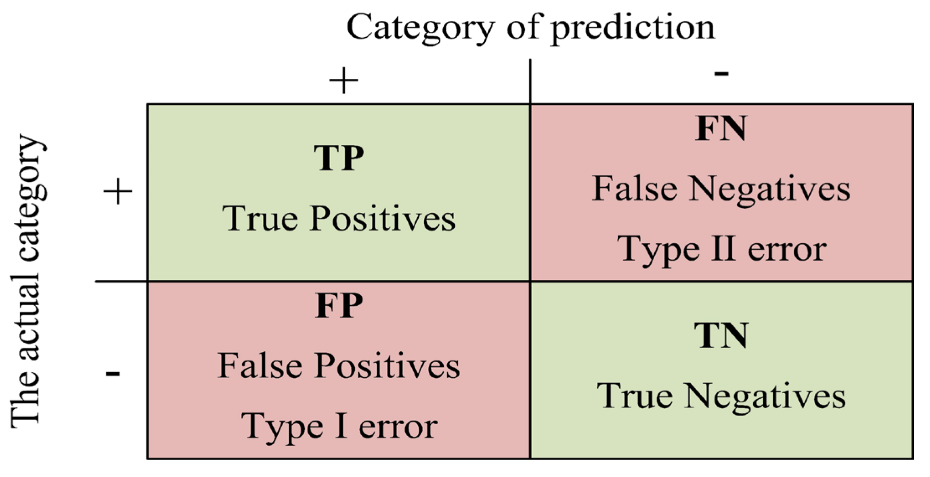


Figure : Structure of the confusion matrix

* **Accuracy**: These performance metrics can be used to assess the effectiveness of this presented system to correctly detect objects from the images and termed as the no. of accurate classifications with the total no. of the classifications (Yan et al., 2021).
* **Precision**: It represents the ratio of the correctly classified positive samples with all positive class predicted classifications.
* **Recall**: It represents the proportion of the correctly forecasted samples with the total no. of the correctly classified positive sample and incorrectly classified negative sample.
* **F1-score**: This performance metric represents harmonic means of recall and precision.
* **Mean average precision (mAP)**: This metric also helps in evaluating and comparing the performance of object recognition (Yue et al., 2022). A curve of Precision-Recall can be derived from the relationship between Recall and Precision.

The average precision refers to the area surrounded by the axes and curves. The value of mAP can be obtained by taking the average of the AP, as illustrated below.

### Experimental results

We have conducted an experiment on the taken experimental set. In this experiment, we place the different objects in from this object recognition system and check when it can detect the object and what is the accuracy of classification.

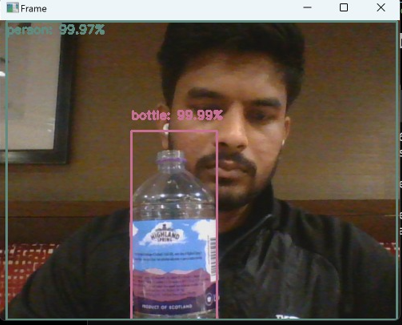


Figure : Detection of bottle object

Figure 25 shows that the proposed system accurately detects the bottle object and creates a bounding box around this object to enable the user to easily notice the detected object (Motwani et al., 2022). This bottle object is detected with about 99.99% classification accuracy. It also detects the person in the background with a classification accuracy of 99.97%.

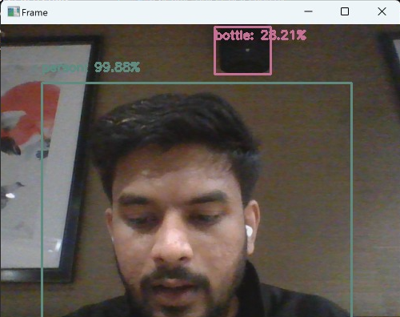


Figure : Detecting a person as a different object

Figure 26 shows that this presented system detects a person with a classification accuracy of 99.88%. This system also detects bottle objects, but it deliberated relatively low classification accuracy of 28.21% due to its small.

### Findings

This project report presented the following key research findings.

* The increased digital transformation also increases the demand for object recognition systems in various applications, such as video surveillance systems, autonomous driving, robot vision, and security and safety applications.
* The existing system is not efficient to accurately and efficiently detect objects from images or videos. Also, they cannot extract the desired information about objects from the images.
* The increased emergence of machine learning and deep learning methods optimize the capabilities of object recognition systems and make them capable of accurately and efficiently recognizing the objects present in the given images or videos.
* The presented object recognition system can correctly localize the objects within the video frames or images and make a bounding box around the recognized objects while measuring the localization accuracy (Sugashini & Balakrishnan, 2023).
* This system can also offer performance insights to recognize diverse categories of objects while addressing which categories are correctly recognized and which possess low accuracy or challenge during detection.
* This system is robust against diverse object variations like occlusion, scale changes, or lighting conditions and can efficiently handle these variations.
* This presented system can be used in different tasks or industries, such as autonomous driving, video surveillance, medical imaging, or retail analytics.

## Testing

The testing of a web application represents a software practice for ensuring that the tested system or solution is functional, works as intended, and meets all the requirements. It also helps to recognize potential errors or bugs at any time or stage of the system development before its release (Cavero-Baptista, 2023). The identification of these bugs during the development phase helps to prevent additional costs to fix them later.

Rigorous testing is essential to create a top-notch web application. So here a no. of efficient and reliable test cases are created to conduct robust testing. For testing the functionality and effectiveness of the proposed web application, the performed testing are functional testing, performance testing, regression testing, and cross-browser testing.

* **Functional testing**: This type of testing helps ensure that the functionalities of the developed web application work as envisioned from the perspectives of the end-user. This testing ensures that the different components of elements of the web application are working well together and appear correctly in the system’s user interface (Palmer, 2022). Here individual testing method is used for testing combinations of different units or elements of code to ensure that they work well together.
* **Regression testing**: This test type also refers to repeated functional testing and it is used to ensure that the functionality of the developed software system is functioning after the modifications in the configurations or codes. When the existing features are improved or new features are developed, then this testing helps in ensuring that the existing features are working as intended. Here different computer-based techniques and tools are used for testing the developed software.
* **Cross-browser testing**: This type of software testing helps in ensuring that the developed web application performs as intended across diverse browsers, for both mobiles and desktops. It compares and analyses the developed web application’s behaviors in diverse browser atmospheres. It helps to pinpoint browser-centric compatibility errors for debugging them quickly and conveniently. It also helps ensure that the developed website delivers an optimized user experience regardless of the type or version of the browser used by the users.
* **Performance testing**: Performance testing helps ensure that the developed web application endures peak user loads and extended activity periods. It represents a process to evaluate how the developed web application performs in terms of stability and responsiveness under a specific workload. This type of testing is typically performed to examine the robustness, speed, and reliability of the application (Gillis, 2023). This testing comprises the performance indicators, such as the response time of the browser, network, & web page, processing time of server request, acceptable simultaneous user volumes, and consumption of processor memory.

## Discussion

Generally, the Haar Cascade algorithm is highly efficient and accurate in recognizing the objects from the input image and videos. However, the accuracy of this detection algorithm depends on several factors, such as quantity and quality of training data, tuning of the parameters of the algorithm, and the detected feature’s complexity. If this algorithm is properly tuned and trained, then it could attain higher detection rates with lower false negatives and positives.

### Average precision of the proposed model

We have considered and compared the YOLOv4 and proposed model on the taken image dataset. The obtained results for recall, precision, and average precision from the conducted experiment for both of the considered models are illustrated in the table. According to this table, the obtained average precision for the proposed model is about 98.678%, which is 6.523% higher as compared to the YOLOv4 model (Wang et al., 2023). The proposed model outperformed YOLOv4 in terms of precision and recall for all the taken indicators.

Table : Obtained results from the conducted experiments on the taken image dataset

|  |  |  |  |
| --- | --- | --- | --- |
| **Evaluation Metrics** | **Evaluation indicators** | **YOLOv4** | **Proposed model** |
| Precision | Person | 95.5% | 99.97% |
| Bottle | 91.2% | 99.99% |
| Recall | Person | 92.9% | 97.95% |
| Bottle | 89.02% | 96.8% |
| Average Precision |  | 92.155% | 98.678% |

Figure : Average precision for the YOLOv4 model

According to Figure 27, the YOLOv4 model demonstrated a precision of 95.5% and 91.2% for person and bottle, respectively, whereas the obtained values of recall for both person and bottle using the YOLOv4 model is 92.9% and 89.02%, respectively. The mean average precision (mAP) for this model is 92.155%.

Figure : Average precision for the proposed model

According to Figure 28, the proposed YOLO-based object recognition model demonstrated a precision of 99.97% and 99.99% for person and bottle, respectively, whereas the obtained values of recall for both person and bottle using the proposed model is 97.95% and 96.8%, respectively. The mean average precision (mAP) for this model is 98.678% (Wang et al., 2023).

### Economic and Commercial feasibility

The presented YOLO-based object recognition system is a very efficient system that can be used by any organization, business, or government firm to accurately detect objects from images for various purposes. The YOLO algorithm provides an ability to enable accurate and quick object recognition for specifically increasing proficiency in different industries, such as retail, manufacturing, and logistics by automating various processes (Ahmad et al., 2020). The businesses also significantly reduce the associated costs with manual object recognition and tracking. This system can offer real-time object or individual recognition and monitoring in the field of video surveillance and security while improving security measures. Further, it could be used to improve quality measures in the production and manufacturing industry along with identifying irregularities to avoid faulty production and reduce associated costs & wastes. Moreover, this system can better manage inventories and minimize the chance of overstock or stockout situations while facilitating an optimized customer service experience. The generated data by this system can be helpful in gaining valuable insights related to product trends, customer behavior, and operational proficiency. The use of this system by the companies demonstrates their commitment to the adoption of cutting-edge technologies, that can attract investors and increase business competitiveness.

# Conclusion

In this part of the project report, I presented the obtained research results, used testing methods, and discussion of the attained research results. In this section, I discussed the performance evaluation metrics, such as precision, recall, and mean average precision. Then I presented the attained experimental results from the conducted experiments. Then I presented the obtained research findings from this research. Then I discussed the testing methods, such as functional testing, regression testing, cross-browser testing, and performance testing. Then I provided a discussion of the obtained experimental results. Finally, I deliberated the economic and commercial feasibility of the presented object recognition system.

Chapter 6: Conclusion and Recommendations

# Conclusion and Recommendations

## Conclusion

This research project successfully developed the intended YOLO-based object recognition system to efficiently and accurately detect different types of objects presented in any image or video. The increased emergence of digital technologies, such as deep learning and machine learning demands for real-time object detection systems to detect any suspicious activity to prevent any criminal activities. The evolution of computer vision provides an ability for computer systems and other devices to extract vital insights from videos and images to better understand the surroundings. The developed YOLO-based object recognition system can outperform other methods and significantly enhance the capability of existing object recognition systems to enhance trust in made predictions, particularly to increase people's security. In this research, I used the YOLO algorithm with the ML and DL methods. Further, a Haar Cascade algorithm to further support the process of object recognition regardless of the scale and location of the image.

This research presented an intelligent object recognition system based on the YOLO algorithm and machine learning methods to better recognize and localize the different types of objects within an image or video frame. This proposed web application system was developed using Python programming language along with the libraries, such as NumPY, TensorFlow, and OpenCV library. The conducted experimental result demonstrated that the presented object recognition system outperforms the other version of YOLO models and it demonstrates an average precision of 98.678%, which is 6.523% higher than the YOLOv4 model.

In this project report, First I focused on introducing this research project for realizing a robust object detection system to accurately perceive and categorize miscellaneous types of objects from the images and videos. Here I provided the research background, aim & objectives, problem statement, research questions, research rationale, description or artifact, and report structure. In the second chapter, I presented a critical literature review of the previously published research papers or journal articles for understanding other authors’ works and used methods, theories, or concepts. In the third chapter, I discussed the chosen mixed research methodology along with the sequential explanatory research design pragmatism research philosophy concurrent embedded research approach. Further, the YOLO algorithm, machine learning methods, Haar Cascade algorithm, product backlog approach, and Agile project management methodology are also discussed. In the fourth chapter, I presented the developed system design along with its implementation. In the fifth chapter, I discussed the obtained research results, testing types, and a discussion of these results. Finally, In the sixth chapter, I concluded this research by summarizing key research findings and providing key recommendations to conduct future work.

## Recommendations

These research findings deliberated the most important and valuable research insights to strengthen the capability of the object recognition process. These research insights can be used as the key directions for exploring this research in the future for more accurate and precise detection of the diverse types of objects from the images and videos. The following recommendations can be considered to proceed with this research study in the future.

* Future research should focus on collecting more representative and diversified images for training purposes to ensure consistent and accurate annotations (Dhanya et al., 2023). Further, the data augmentation methods can be helpful in increasing the size and variety of the dataset.
* Future research should utilize the latest versions of the YOLO algorithm and apply transfer learning to train the models on the larger dataset for fostering better convergence, particularly for the restricted training data.
* The non-maximum suppression can be further optimized to better eliminate the redundant overlying bounding boxes for further refining the detections while incorporating the contextual information to improve detection accuracy and decrease false positives.
* The researchers can use hardware acceleration methods, such as TPU or GPU to boost the inference procedure to specifically enhance the real-time performance measures of this developed object recognition system.
* The research should focus on utilizing diverse hyperparameter settings for realizing an optimized configuration according to the task and dataset. Further, hyperparameter optimization methods, such as random search and grid search could be utilized for automating the detection process (Ramík et al., 2013).
* The researchers should focus on using ensemble methods by combining multiple YOLO algorithms to improve the accuracy and robustness of the object recognition system.

# References

Ahmad, T. *et al.* (2020) ‘Object detection through modified yolo neural network’, *Scientific Programming*, 2020, pp. 1–10. doi:10.1155/2020/8403262.

Bell, R., Warren, V. and Schmidt, R. (2022) The application of concurrent or sequential mixed-methods research designs and their methodological implications: Investigating tacit knowledge, its use, and application in Automotive Development, Sage Research Methods. Available at: https://methods.sagepub.com/case/concurrent-sequential-mixed-methods-research-methodological-implications (Accessed: 08 January 2024).

Boesch, G. (2024) Object detection in 2024: The definitive guide, viso.ai. Available at: https://viso.ai/deep-learning/object-detection/ (Accessed: 08 January 2024).

Cavero-Baptista, L. (2023) *Web application testing: The basics of web app test automation*, *Leapwork*. Available at: https://www.leapwork.com/blog/web-application-testing-the-basics-of-web-app-test-automation#:~:text=Web%20testing%2C%20or%20web%20application,day%2Dto%2Dday%20basis (Accessed: 08 January 2024).

Cho, H. *et al.* (2014) ‘A multi-sensor fusion system for moving object detection and tracking in urban driving environments’, *2014 IEEE International Conference on Robotics and Automation (ICRA)* [Preprint]. doi:10.1109/icra.2014.6907100.

Contributor, T. (2015) *What is object recognition?: Definition from TechTarget*, *WhatIs*. Available at: https://www.techtarget.com/whatis/definition/object-recognition#:~:text=Object%20recognition%20allows%20robots%20and,of%20appearances%20from%20different%20angles. (Accessed: 08 January 2024).

Dhanya, K.A. *et al.* (2023) ‘Detection of network attacks using machine learning and Deep Learning Models’, *Procedia Computer Science*, 218, pp. 57–66. doi:10.1016/j.procs.2022.12.401.

Ehlers, A. (2017) *Object detection using feature mining in a distributed machine learning framework* [Preprint]. doi:10.51202/9783186855107.

Gallagher, J. (2023) *What is object detection? the ultimate guide.*, *Roboflow Blog*. Available at: https://blog.roboflow.com/object-detection/ (Accessed: 08 January 2024).

George, T. (2023) Mixed methods research: Definition, guide & examples, Scribbr. Available at: https://www.scribbr.com/methodology/mixed-methods-research/ (Accessed: 08 January 2024).

Gillis, A.S. (2023) *What is performance testing and how does it work? – TechTarget definition*, *Software Quality*. Available at: https://www.techtarget.com/searchsoftwarequality/definition/performance-testing (Accessed: 08 January 2024).

Girshick, R. *et al.* (2016) ‘Region-based convolutional networks for accurate object detection and segmentation’, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1), pp. 142–158. doi:10.1109/tpami.2015.2437384.

Gupta, A. (2022) Some amazing applications of opencv library, Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2021/09/some-amazing-applications-of-opencv-library/#:~:text=OpenCV%20is%20a%20great%20tool,landmark%20detection%2C%20and%20much%20more. (Accessed: 08 January 2024).

Hashim, S. and Mccullagh, P. (2023) ‘Face detection by using Haar Cascade classifier’, Wasit Journal of Computer and Mathematics Science, 2(1), pp. 1–5. doi:10.31185/wjcm.109.

Hu, L. and Li, Y. (2021) ‘Micro-YOLO: Exploring efficient methods to compress CNN based object detection model’, *Proceedings of the 13th International Conference on Agents and Artificial Intelligence* [Preprint]. doi:10.5220/0010234401510158.

Ivankova, N.V., Creswell, J.W. and Stick, S.L. (2006) ‘Using mixed-methods sequential explanatory design: From theory to practice’, Field Methods, 18(1), pp. 3–20. doi:10.1177/1525822x05282260.

Jaiswal, A. (2023) Guide to haar cascade algorithm with object detection example, Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2022/04/object-detection-using-haar-cascade-opencv/ (Accessed: 08 January 2024).

Jia, P. *et al.* (2022) ‘PV-YOLO: An object detection model for panoramic video based on yolov4’, *2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML)* [Preprint]. doi:10.1109/cacml55074.2022.00018.

Kang, K. *et al.* (2018) ‘T-CNN: Tubelets with Convolutional Neural Networks for object detection from videos’, *IEEE Transactions on Circuits and Systems for Video Technology*, 28(10), pp. 2896–2907. doi:10.1109/tcsvt.2017.2736553.

Kristian, Y., Salim, H.A. and Setyati, E. (2020) ‘Detection of militia object in Libya by using Yolo Transfer Learning’, *Jurnal Teknologi dan Manajemen Informatika*, 6(1). doi:10.26905/jtmi.v6i1.4025.

Kumari, K. (2023) *A basic introduction to object detection*, *Analytics Vidhya*. Available at: https://www.analyticsvidhya.com/blog/2022/03/a-basic-introduction-to-object-detection/ (Accessed: 08 January 2024).

Kundu, R. (2023) Yolo algorithm for object detection explained [+examples], YOLO Algorithm for Object Detection Explained [+Examples]. Available at: https://www.v7labs.com/blog/yolo-object-detection (Accessed: 08 January 2024).

Laoyan, S. (2022) What is agile methodology? (a beginner’s guide) [2023] • asana, Asana. Available at: https://asana.com/resources/agile-methodology (Accessed: 08 January 2024).

Liu, Y. et al. (2021) ‘A survey and performance evaluation of deep learning methods for small object detection’, Expert Systems with Applications, 172, p. 114602. doi:10.1016/j.eswa.2021.114602.

Motwani, N.P., S, S. and Singh, U. (2022) ‘Object detection and tracking for autonomous vehicles using deep learning technique- Yolo’, *2022 International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON)* [Preprint]. doi:10.1109/smartgencon56628.2022.10083703.

Niemeyer, J., Rottensteiner, F. and Soergel, U. (2014) ‘Contextual classification of LIDAR data and building object detection in urban areas’, *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, pp. 152–165. doi:10.1016/j.isprsjprs.2013.11.001.

Palmer, J. (2022) *Web application testing: Getting started with functional testing: Rainforest qa*, *Rainforest QA Blog | Software Testing Guides*. Available at: https://www.rainforestqa.com/blog/web-application-testing (Accessed: 08 January 2024).

Patel, R. (2023) *Object detection in computer vision*, *Plutomen*. Available at: https://pluto-men.com/object-detection-in-computer-vision/ (Accessed: 08 January 2024).

Ramík, D.M. *et al.* (2013) ‘A machine learning based intelligent vision system for Autonomous Object Detection And Recognition’, *Applied Intelligence*, 40(2), pp. 358–375. doi:10.1007/s10489-013-0461-5.

Sangeetha, T. *et al.* (2023) ‘Face mask detection and social distancing using machine learning with Haar Cascade algorithm’, *Advances in Intelligent Systems and Computing*, pp. 987–999. doi:10.1007/978-981-19-5443-6\_72.

Shamkuwar, S. *et al.* (2022) ‘Machine learning based theft detection using Yolo Object Detection’, *International Journal of Scientific Research in Science and Technology*, pp. 267–271. doi:10.32628/ijsrst229241.

Simplilearn (2023) *What is Computer Vision: Applications, benefits and how to learn it*, *Simplilearn.com*. Available at: https://www.simplilearn.com/computer-vision-article (Accessed: 08 January 2024).

Singh, H. (2019) ‘Image processing using machine learning’, *Practical Machine Learning and Image Processing*, pp. 89–132. doi:10.1007/978-1-4842-4149-3\_5.

Situ, Z. et al. (2023) ‘A transfer learning-based YOLO network for sewer defect detection in comparison to classic object detection methods’, Developments in the Built Environment, 15, p. 100191. doi:10.1016/j.dibe.2023.100191.

Somani, A. et al. (2023) ‘A survey paper on live criminal detection using facial recognition and tracking algorithms’, International Research Journal of Modernization in Engineering Technology and Science, 5(5). doi:10.56726/irjmets39177.

Srivani, K.D. (2023) Yolo algorithm for Custom Object Detection, Analytics Vidhya. Available at: https://www.analyticsvidhya.com/blog/2022/06/yolo-algorithm-for-custom-object-detection/#:~:text=YOLO%20stands%20for%20You%20Only,detection%20algorithms%20available%20so%20far. (Accessed: 08 January 2024).

Sugashini, T. and Balakrishnan, G. (2023) ‘Yolo Glass: Video-based smart object detection using squeeze and Attention Yolo Network’, *Signal, Image and Video Processing* [Preprint]. doi:10.1007/s11760-023-02855-x.

SuperAnnotate (2021) *Introduction to computer vision: History and applications*, *SuperAnnotate*. Available at: https://www.superannotate.com/blog/introduction-to-computer-vision (Accessed: 08 January 2024).

Tao, J. et al. (2017) ‘An object detection system based on Yolo in traffic scene’, 2017 6th International Conference on Computer Science and Network Technology (ICCSNT) [Preprint]. doi:10.1109/iccsnt.2017.8343709.

Verschae, R. and Ruiz-del-Solar, J. (2015) ‘Object detection: Current and future directions’, *Frontiers in Robotics and AI*, 2. doi:10.3389/frobt.2015.00029.

Wang, R.J., Li, X. and Ling, C.X. (2019) *Pelee: A Real-Time Object Detection System on Mobile Devices* [Preprint]. doi: https://doi.org/10.48550/arXiv.1804.06882.

Wang, Y. *et al.* (2023) ‘VV-Yolo: A vehicle view object detection model based on improved Yolov4’, *Sensors*, 23(7), p. 3385. doi:10.3390/s23073385.

Yan, W., Liu, T. and Fu, Y. (2021) ‘Yolo-tight: An efficient dynamic compression method for Yolo Object Detection Networks’, *2021 13th International Conference on Machine Learning and Computing* [Preprint]. doi:10.1145/3457682.3457740.

Yegulalp, S. (2024) What is tensorflow? The Machine Learning Library explained, InfoWorld. Available at: https://www.infoworld.com/article/3278008/what-is-tensorflow-the-machine-learning-library-explained.html (Accessed: 08 January 2024).

Yue, X. *et al.* (2022) ‘Yolo-GD: A deep learning-based object detection algorithm for empty-dish recycling robots’, *Machines*, 10(5), p. 294. doi:10.3390/machines10050294.

Yu, J. *et al.* (2016) ‘UnitBox: An Advanced Object Detection Network’, *Proceedings of the 24th ACM international conference on Multimedia* [Preprint]. doi:10.1145/2964284.2967274.

Zhu, Y., Zhu, J. and Zhang, R. (2014) ‘Contextual object detection with spatial context prototypes’, *IEEE Transactions on Multimedia*, pp. 1–1. doi:10.1109/tmm.2014.2321534.