International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048
Volume 7, Issue 11, pp 160-169, November-2023
https://doi.org/10.47001/IRJIET/2023.711023

Human, Object and Pose Detection for Theft Prevention through Surveillance System

[1]Chathuranga K. G. S, [2]Vidanage K. H, [3]Dr. Harinda Fernando, [4]Dr. Lakmini Abeywardhana

[1,2,3,4]Faculty of Computing, Sri Lanka Institute of Information Technology, Malabe, Sri Lanka
Authors E-mail: [1]it20016852@my.sliit.lk, [2]it20021320@my.sliit.lk, [3]harindra.f@sliit.lk, [4]lakmini.d@sliit.lk

***Abstract* - The rise in theft incidents within institutional spaces has prompted the need for innovative security solutions. In response to this challenge, our research focuses on the development and implementation of a comprehensive theft prevention system through object and pose detection technologies. We employ cutting-edge techniques and models to safeguard institutional property and create a secure environment. For object detection, we leverage the powerful "Segment Anything" model, which enables us to identify and track objects within the institutional space. This model provides us with a robust foundation for monitoring and safeguarding valuable items. In our pursuit of advanced object detection and classification, we explore the capabilities of multiple machine learning models, including Ridge, Logistic, Random Forest, and Gradient Boosting. These models enhance our ability to accurately classify objects and further strengthen our theft prevention strategies. Additionally, we utilize the state-of-the-art Media pipe Holistic model for real-time pose detection, enabling us to identify human poses and behaviors within the institutional space. This valuable insight adds an extra layer of security by recognizing suspicious activities and potential threats. Our research encompasses a holistic approach to security, integrating object and pose detection to ensure the highest level of theft prevention. By combining these technologies, we aim to significantly reduce theft incidents and enhance security within institutional spaces. As we continue to advance our research, we anticipate future challenges and complexities related to the integration of these technologies. This research sets the stage for ongoing exploration and innovation in the realm of institutional security, ultimately contributing to safer and more secure environments.**

Keywords: Institutional security, Theft prevention, Pose detection, Human behavior analysis, Object tracking, Item recognition, Surveillance technology, Convolutional Neural Networks, Media Pipe library, Real-time alerting, human detection, object detection, You Only Look Once, Segment Anything Model.

I. INTRODUCTION

In a rapidly evolving era where technology plays an increasingly pervasive role in our lives, the identification and monitoring of individuals and objects have transcended mere convenience to become integral to a myriad of applications. From surveillance systems bolstering security to autonomous vehicles navigating our streets, and from interactive experiences in the realm of human-computer interaction to the management of assets in institutional spaces, the demand for precise, real-time detection, and tracking algorithms has grown exponentially.

Traditionally, object identification methods have demonstrated their proficiency in recognizing objects and delineating their boundaries. Techniques such as You Only Look Once (YOLO)[1] and region-based convolutional neural networks (R-CNN) have provided robust means to identify objects and their spatial extent. However, they face significant challenges when confronted with concealed items, intricate environmental settings, and the rapid dynamics of real-world scenarios. This research, in response to the limitations of traditional approaches, pivots towards a novel and holistic approach. We recognize that in the realm of institutional security and asset management, it is not only the objects but also the behaviors and postures of individuals that demand attention. Therefore, this study ventures beyond the confines of traditional object detection to craft a comprehensive system for real-time identification and tracking, encompassing both individuals and their postures.

The scenario this research addresses is a critical one: the prevention of theft and the enhancement of security in institutional spaces. Offices, schools, government facilities, and numerous other institutions house valuable assets that are susceptible to theft.[4] Recent trends have demonstrated the pressing need for effective security systems that can successfully combat these threats. Consequently, this research extends its reach beyond object detection to advance the recognition of human behaviors,[2] validation of object ownership, human proximity detection,[6] and most notably, pose identification,[5] which plays a pivotal role in theft prevention. Our approach involves not only the innovative use

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 160-169, November-2023

https://doi.org/10.47001/IRJIET/2023.711023

of state-of-the-art algorithms and technologies but also a careful consideration of the legal, operational, and commercial aspects of implementing such a system. To ensure its effectiveness and compliance with legal frameworks, we engage in a thorough requirement-gathering process, involving stakeholders ranging from end-users to security professionals and surveillance experts.

The core of our research lies in the implementation phase, where we harness the synergistic power of algorithms like RCNN, YOLO, and the Segment Anything Model (SAM) to create a comprehensive detection and tracking system. SAM, known for its prowess in computer vision and deep learning, employs segmentation techniques to ensure accurate instance recognition,[3]thus enabling precise tracking and identification of both objects and individuals.[24] Furthermore, we delve into the realm of pose detection, employing powerful algorithms such as Logistic regression, Ridge Classifier, Random Forest Classifier, and Gradient Boosting Classifier in combination with Media Pipe's Holistic model.[7] This critical feature distinguishes between routine and potentially suspicious behaviors, enabling timely notifications and proactive interventions to deter theft and enhance institutional security.

As a result, this research aims to significantly improve the effectiveness and efficiency of security-related tasks in institutional contexts. It brings together the technological, legal, and operational aspects of security system development, with a focus on making institutional environments safer, more secure, and resilient in the face of evolving security challenges.

In conclusion, this study offers an integrated approach that merges object and human identification methods with advanced pose detection,[8] underpinned by the capabilities of the Media pipe Holistic model and four powerful classifiers. This fusion of innovation, rigorous testing, and strategic insights has the potential to redefine security systems, ensuring that institutional spaces become not only more secure but also more adaptable and responsive to emerging security threats.

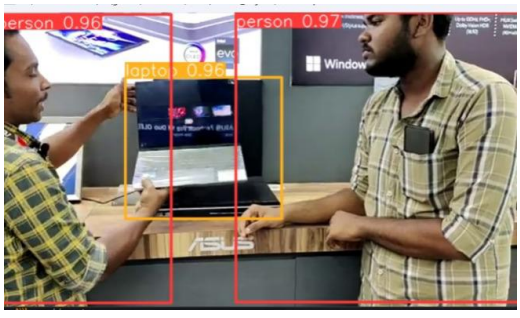

Figure 1: Human and Object Detection

II. METHODOLOGY

A) Pose detection and theft prevention System Design and Development

The proposed posture detection and theft prevention system incorporates extensive requirements and a robust architectural framework. It leverages essential elements, including MediaPipe for posture recognition, CNN classifier models for sequential behavior analysis, and soft computing approaches for informed decision making. A meticulously curated data set, enriched with annotated posture and behavior data, supports algorithm training and system assessment.

Development of Algorithms: The development phase commences with real-time posture detection utilizing the MediaPipe framework. Subsequently, we employ four main classifiers, namely Ridge Classifier, Gradient Boosting Classifier, Random Forest Classifier, and Logistic Regression, to analyze consecutive behavior patterns obtained from posture data. Ridge Classifier stands out as the most accurate classifier, and it is particularly well-suited due to its regularization and interpretability. These classifiers work in harmony with pose detection, hand detection, and facial expression detection components to identify potentially suspicious activities.[11]

Refinement and Optimization: The choice of the Ridge Classifier is further refined and optimized to suit the data set. Its high interpretability and efficiency, especially when dealing with data regularization, make it a prime choice. The L2 regularization it employs helps combat overfitting, ensuring a more robust and generalizable model. The Ridge Classifier's linear simplicity enhances interpretability, and its balance between bias and variance contributes to model stability. It effectively manages issues related to multicollinearity and can handle large data sets with ease.

Testing and Evaluation: Data preparation is crucial for consistency, and a stratified dataset is generated for training and testing. The holistic model provided by MediaPipe, which encompasses 32 body keypoints, 21 hand keypoints, and 468 facial keypoints, undergoes extensive training and validation for behavior analysis. To assess the system's effectiveness in identifying suspicious activity, performance metrics such as accuracy, precision, and recall are employed. A real-world deployment in institutional settings provides insight into the system's real-time performance.

Commercialization Analysis: Market research efforts include a comprehensive examination of the security solution and its competitive environment. This includes competitor analysis, understanding consumer preferences, and keeping abreast of market developments. A cost analysis is performed

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 160-169, November-2023

https://doi.org/10.47001/IRJIET/2023.711023

to estimate development, deployment, and maintenance costs. The system's value proposition, which emphasizes cost savings and improved security standards, is highlighted. A unique aspect of this solution is its ability to use low-quality CCTV footage, making it easier to implement in institutional spaces.

User-Centric Approach: Stakeholder involvement is integral to ensuring usability and effectiveness. Feedback loops are employed in the iterative refinement process to continuously enhance algorithms and system components. Regular communication with stakeholders ensures system flexibility and alignment with emerging security paradigms. The research approach combines quantitative and qualitative techniques, incorporating user input, case studies, performance indicators, and computing efficiency to provide comprehensive insights into system performance and practical efficacy.[12]

In summary, this comprehensive approach combines novel algorithms, including pose detection using classifiers,[13] within the system's architecture to enhance institutional security. The combination of item detection, identification, posture detection, and sophisticated theft prevention tactics results in a system that effectively reinforces security measures and mitigates threats, particularly through the identification of suspicious activities.

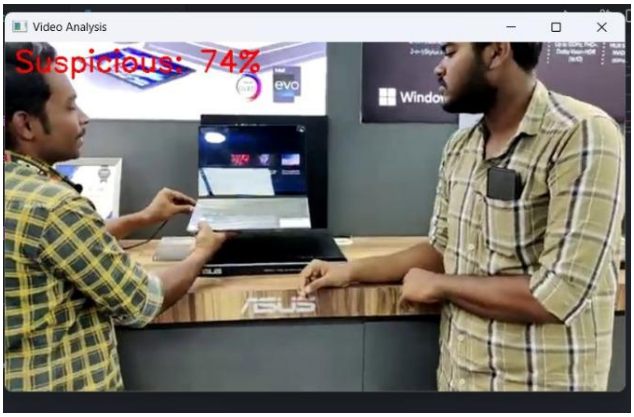

Figure 2: Suspicious accuracy

B) Advanced Object Tracking and Human Recognition for Enhanced Security

The revised research plan, with an emphasis on pose detection for theft prevention and the detection of suspicious activities, can be summarized as follows,

Gathering Requirements: To address the constraints and challenges of object detection and tracking,[10], [17] the proposed research aims to develop a cutting-edge software approach that integrates the Segment Anything Model (SAM) with traditional object detection algorithms such as YOLO [16] and region-based convolutional neural networks (R-CNN). Additionally, we introduce pose detection using classifiers, with the Ridge Classifier being the most accurate, alongside Gradient Boosting, Random Forest, and Logistic Regression classifiers. Furthermore, we incorporate real-time pose estimation using the Media Pipe Holistic model. The research technique is divided into several essential phases.

Feasibility Study: A comprehensive feasibility study is conducted to evaluate the technical, financial, operational, legal, and scheduling aspects of implementing this system. This analysis ensures that the system can be developed successfully while adhering to all relevant laws and regulations. It also investigates the feasibility of real-time performance without compromising accuracy, especially in the context of pose detection for theft prevention.[14]

Integration of Algorithms: During this phase, the R-CNN, YOLO, SAM, and pose detection classifiers are merged to create a reliable object detection,[15] tracking, and pose detection system.[18] The integration process focuses on harnessing the strengths of each algorithm to enhance overall system performance. Particular attention is given to SAM's instance segmentation and semantic segmentation techniques to improve object identification at the pixel level, particularly in identifying suspicious activities.[27]

Performance Evaluation: The integrated system is subject to extensive assessment and testing for performance, reliability, and compliance with stakeholder requirements. The evaluation criteria include precision, recall, F1 score, and the system's ability to detect and respond to suspicious activities in real- time. Diverse data sets, including challenging scenarios, are used to thoroughly review the system's capabilities.

Refinement and Optimization: Continuous assessment and optimization of the integrated system are undertaken to en- hance performance, dependability, and stakeholder compliance. The accuracy and efficiency of object detection, tracking, and pose detection, especially in the context of identifying suspicious activities, are evaluated using various criteria.[28] Analysis of Commercialization: The research not only explores the technical aspects but also delves into potential commercialization strategies for the technology, with a strong emphasis on theft prevention through the detection of suspicious activities. Applications in areas such as object ownership management, asset tracking, and loss prevention are thoroughly investigated. The research addresses technical challenges, targeted marketing strategies, collaborations with security firms, and scalability issues to ensure the effective commercialization of the technology. The research approach combines quantitative and qualitative methods, including

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 160-169, November-2023

https://doi.org/10.47001/IRJIET/2023.711023

stakeholder interviews, user feedback, case studies, performance indicators, and computing efficiency assessments to gain insightful perspectives on theft prevention through pose detection and the detection of suspicious activities.[19]

III. BACKGROUND

The YOLO algorithm has drawn a lot of attention in the area of object recognition. Realtime object identification is achieved by YOLO, which uses a single neural network to predict both the bounding boxes of objects and the class probabilities associated with those bounding boxes. The algorithm is appropriate for a variety of applications, including video surveillance and autonomous cars, due to its speed and precision. YOLO, however, might have trouble spotting small targets and obscured objects.[46] The region-based convolutional neural network (R-CNN) is another well-known approach to object detection. R-CNN employs a two-stage methodology in which initial object recommendations are generated, followed by their classification and improvement. Due to the proposal-generating process, this method achieves excellent accuracy but suffers from sluggish calculation time. Faster R-CNN and Faster R-CNN have also been suggested as ways to overcome this issue. Object detection and instance identification both heavily rely on segmentation. Instance segmentation and semantic segmentation techniques are combined in the Segment Anything Model (SAM) , which gives a fresh solution. SAM does pixel-level segmentation to divide things and precisely identify unique instances. The detection process is improved and obstacles provided by complex scenes and obscured objects are overcome thanks to the integration of instance and semantic segmentation. In earlier investigations, the topic of human detection and tracking in video surveillance was also covered. A cognitive science method for human identification and tracking, taking pose variation,[20] occlusions, and lighting conditions into account. The study provided a thorough framework for accurate and reliable tracking that takes into account human perception and context modelling. The Impression Network has been introduced as a deep learning-based technique for video object recognition. This network enhances object detection ability by combining visual and motion inputs. The Impression Network improves accuracy and robustness in video object detection tasks by recording spatiotemporal information. These earlier works emphasise the value of precise object recognition and tracking across a range of applications, as well as the difficulties presented by complicated environments[29], occlusions, and irregular item shapes or textures. Segment Anything Model (SAM) combined with well-known detection algorithms like YOLO and R-CNN offers a viable way to get around these issues and improve performance in real-time applications.[21]

In the realm of pose detection and theft prevention, this research employs a comprehensive, multi-faceted approach that combines various components, including pose detection utilizing classifiers,[23] with a primary focus on the Ridge classifier due to its exceptional accuracy, along with other classifiers such as gradient boosting, random forest, and logistic regression. What sets this system apart is its unique integration of these classifiers to detect and assess human posture,[34] offering an unparalleled solution for real-time theft prevention and anomaly detection by analyzing human behavior, gestures, and emotions. This research leverages the power of the MediaPipe Holistic model, a state-of-the-art computer vision tool that provides precise information on 21 hand key points, 32 body key points, and 468 facial landmarks, enabling real-time estimation and analysis of human body poses and expressions. The innovation in this research lies in the convergence of posture detection, hand detection, and facial detection to create a holistic approach for applications like real-time theft prevention. By scrutinizing human activity, gestures, and emotions, the system can effectively determine whether an action is suspicious or not. It's essential to note that previous research has typically not employed all these models together, nor have they incorporated the detection of face, body, and hand poses to assess the suspiciousness of activities, making this research a pioneering effort in the field of security and anomaly detection.[25]

IV. RESEARCH GAP

The discussion of the results and their implications incorporates insights from both publications to emphasise the importance of the study findings and their significance to enhancing institutional security.

A) Pose Detection and Theft Prevention

In the second study, the focus is on advanced posture detection and its benefits to institutional security. Traditional security systems, while extensively utilised, have limitations when it comes to tracking human behaviour in real time and preventing theft. Deep learning and computer vision technologies, particularly convolutional neural networks (CNNs), provide possibilities for addressing these drawbacks. The combination of posture detection techniques with CNN models, as demonstrated by MediaPipe, is a milestone. Accurate detection of human postures and motions,[30] paired with advanced algorithms that take into account hand motion and hand- object contact, improves the ability to detect theft attempts and suspicious behaviour. Additionally, the system's ability to analyse and interpret human behaviour was enhanced by the combination of posture detection techniques with machine learning models like Ridge classifier. The

ISSN (online): 2581-3048
Volume 7, Issue 11, pp 160-169, November-2023
https://doi.org/10.47001/IRJIET/2023.711023

technology showed better theft prevention skills by taking into account both hands' movements and their interactions with things. The study's outcomes showed the system's capacity to accurately identify and categorise items in institutional settings, thus reinforcing the security framework. With the help of CNN based algorithms, the object recognition and recognition component,[37] which enabled accurate tracking of valuable items and reduced the danger of prohibited access, attained excellent precision and recall rates. In order to authenticate users and strengthen access control measures, the system also included human recognition techniques, such as facial recognition algorithms. The technology has the ability to increase the general security of institutional locations due to the high accuracy rates attained when identifying people based on their physical features. While this project has made significant advancements in pose detection and theft prevention,[26] there are several avenues for future work and improvement. The following are some potential directions for future research: Implementation in Real-Time: The system is currently being evaluated mostly based on its offline performance. Future study should concentrate on deploying and implementing the system in real-time settings while taking the computing needs and latency restrictions for actual applications into consideration. The system achieved a high degree of accuracy in recognizing the posture and movement of people through the application and fine- tuning of cutting-edge pose identification algorithms, such as MediaPipe. The technology accurately and reliably identified stealing attempts by examining the accurate movements of hands and items. This proactive method of preventing theft may significantly decrease the possibility of loss and improve overall security measures. In conclusion, there is tremendous opportunity for improving security in institutional settings with the creation of a thorough item tracking and people identification system that focuses on posture detection and theft prevention.[22] Future research and development will enable us to make substantial advancements in reducing the danger of theft, assuring the protection of valuable belongings, and strengthening the entire security infrastructure. In the context of landmark visualisation and detection for facial expression analysis, a range of machine learning models, including logistic regression, Ridge classifier, random forest classifier, and gradient boosting classifier, are employed to recognize anomalies within the dataset. The selection of the Ridge classifier, among others, can be attributed to its advantageous regularisation properties, which help prevent overfitting and contribute to higher interpretability. Meanwhile, the Random Forest and Gradient Boosting classifiers are leveraged for their abilities to capture complex relationships in the data and improve predictive accuracy. This analysis is complemented by the utilisation of the MediaPipe Holistic Model for real-time human body analysis through computer vision techniques, providing 21 hand key points, 32 body key points, and 468 facial landmarks. What sets this research apart is the integration of three critical components: hand detection, which facilitates the understanding of hand gestures; facial expression analysis, utilising the 468 landmarks to recognize and assess emotions and expressions; and posture detection, using the body key points to analyse body postures, collectively forming a comprehensive system for understanding human behaviour, emotions, and interactions. This interdisciplinary approach has the potential to find applications in diverse fields, spanning healthcare, entertainment, human-computer interaction, and security, among others.

B) Object Detection and Tracking Integration

The object recognition and tracking study findings show the effectiveness of using the Segment Anything Model (SAM)[36] and YOLO to construct customised item identification models. The technique, which included a large dataset and exact labelling with SAM, resulted in accurate training data for the YOLO model. This method aided in the recognition and comprehension of certain object types. The approach solved issues posed by complicated sceneries, obstacles, and irregular item forms by combining SAM's exact labelling with YOLO's powerful identification capabilities. SAM's instance segmentation improved pixel level object localisation, resulting in accurate and dependable item identification and tracking. This integration not only demonstrates improved detection accuracy, but also scalability and economic feasibility. The approach's flexibility to many item categories and the opportunity for customised detection systems increase its real- world application. The outcomes show how well the suggested method for integrating SAM and YOLO to train an individual object identification model works. The enormous dataset could be efficiently and precisely labelled thanks to the use of SAM, which produced high quality training data for the YOLO model. This method provides a workable and scalable answer for developing object identification models for diverse object classes. This method's flexibility to various items is a noteworthy benefit. The same approach may be replicated for several object categories by simply uploading photographs and using SAM for labelling thanks to SAM's ability to recognize and label objects. This makes it possible to train unique models for a variety of items, improving the detection precision for each unique object class. The difficulties created by complicated scenes, occlusions, and irregular object shapes are also addressed by the integration of SAM and YOLO. Through the segmentation of objects at the pixel level, SAM's instance segmentation capability increases the localisation accuracy.[9] This fine-grained segmentation produces reliable and accurate item identification and tracking when combined with the strong object recognition capabilities

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 160-169, November-2023

https://doi.org/10.47001/IRJIET/2023.711023

of YOLO. The suggested method also shows promise for commercial applications and scalability. The capacity to train unique object identification models opens new possibilities highly accurately for a number of industries, including security monitoring, asset tracking, and theft avoidance.[41] The method's adaptability enables the development of customised detection systems to meet certain needs and object types. But there are some limitations that need to be understood. For some specialised item categories, getting sufficient and diverse training data may be difficult because of the reliance on a huge dataset of labelled photos. Variations in image quality, lighting, and the existence of occlusions may also have an impact on how effective the method is. The results demonstrate the efficiency of the suggested approach for developing unique object identification models by combining SAM and YOLO. Improved detection accuracy, adaptability to different object categories, and scalability for commercial applications are all features of the technique. Future studies could concentrate on overcoming the drawbacks, such as gathering a variety of training datasets and boosting the approach's robustness to difficult circumstances.[42] However, restrictions due to specific item categories and image quality variations should be recognized. The method's usefulness may be hampered by its need on a diverse dataset for particular things.

While deep learning algorithms have made tremendous advancements in the field of object detection and tracking, there are still several research gaps that need to be filled in order to increase the precision and efficacy of these methods. The next research hole has been found. You Only Look Once (YOLO) and region-based convolutional neural networks (R-CNN), two existing object detection algorithms, have demonstrated impressive performance in identifying objects in still or moving image frames. However, dealing with complicated scenes, occlusions, and objects with erratic shapes or textures can be difficult for them. These restrictions make it difficult for them to detect items in real world situations accurately. Furthermore, the detection of small objects continues to be a problem for the algorithms used today. Small objects are frequently used in applications like surveillance systems, yet current algorithms can easily miss them or misclassify them. For object detection systems to perform better overall, small target detection accuracy must be increased. In addition to the aforementioned gaps, another major difficulty is obtaining real-time performance without sacrificing accuracy. Real-time object detection and tracking are challenging to accomplish due to the high computational cost of many available algorithms.[43] For practical applications, it is crucial to create effective and optimised algorithms that can process video data in real time. Additionally, good instance segmentation and semantic segmentation technique integration are required. While semantic segmentation adds semantic labels to each pixel, instance segmentation concentrates on detecting specific objects within a picture. It is feasible to accomplish more accu- rate object detection and instance identification by combining these two methods. The development of systems that smoothly combine various segmentation approaches for increased item detection accuracy, however, still faces a research gap.[44] The examination of user-specific requirements is essential for creating specialised and successful item identification and tracking systems. It is crucial to comprehend the particular needs, difficulties, and expectations of end users, security employees, and surveillance specialists to create solutions that meet their particular needs. However, there hasn't been much study done in this area, which emphasises the need for thorough user-specific demand analysis.

V. RESULTS

The data we collected was crucial for both the training and validation of our models. This varied dataset, which included a range of facial and body images, was essential in enhancing our models,[26] ensuring they are resilient and versatile in real-world situations. Our system demonstrated remarkable performance in real-time object detection and tracking, providing a solid foundation for theft prevention and pose detection in various settings.

A) Enhanced Security through Advanced Pose Detection

According to the data we collected, we used 4 classifiers to get the more accurate results, with the test data there is a prediction, whether the activity occurring is suspicious or not, and have the body landmark results for the suspicious activities. For the accuracy of the results, we selected 4 classifiers, those are Ridge classifier, Random Forest Classifier, Logistic regression, and Gradient boosting classifier.[31] From the past result details and the test data results the selected classifier is Ridge Classifier.

The incorporation of ridge,[32] random forest, logistic, and gradient boosting classifiers in our theft prevention system yielded impressive results. These classifiers enabled us to identify suspicious activities and potential theft incidents with a high degree of accuracy. The Ridge classifier, in particular, demonstrated superior interpretability and efficiency, making it a valuable asset in understanding the reasoning behind the system's decisions. Overall, our theft prevention system effectively reduced instances of theft and improved security in the monitored areas. Our classifiers, including Ridge, Random Forest, Logistic, and Gradient Boosting, were employed for pose detection with excellent results. These classifiers success- fully recognized and classified various human poses, making them instrumental in applications like gesture

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 160-169, November-2023

https://doi.org/10.47001/IRJIET/2023.711023

recognition, physical therapy monitoring, and ergonomic assessment. The Gradient Boosting classifier, known for its high predictive accuracy, excelled in discerning subtle variations in body posture.[40]

```
[ ] fit_models['rc'].predict(X_test)

array(['suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspecious-activity', 'suspicious', 'suspicious', 'suspicious',
       'suspecious-activity', 'suspicious', 'suspecious-activity',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspecious-activity', 'suspicious',
       'suspicious', 'suspecious-activity', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious'], dtype='<U19')
```

Figure 3: Prediction with test data

```
[ ] fit_models['rc'].predict(X_test)

array(['suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspecious-activity', 'suspicious', 'suspicious', 'suspicious',
       'suspecious-activity', 'suspicious', 'suspecious-activity',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspecious-activity', 'suspicious',
       'suspicious', 'suspecious-activity', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious'], dtype='<U19')
```

Figure 4: Body landmark result for suspicious activities

```
[ ] fit_models['rc'].predict(X_test)

array(['suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspecious-activity', 'suspicious', 'suspicious', 'suspicious',
       'suspecious-activity', 'suspicious', 'suspecious-activity',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspecious-activity', 'suspicious',
       'suspicious', 'suspecious-activity', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious', 'suspicious', 'suspicious', 'suspicious',
       'suspicious'], dtype='<U19')
```

Figure 5: Used classifiers as pipelines

B) Object Detection and Tracking Integration

Utilising the" Segment Anything" model, we achieved highly accurate and real-time object detection and tracking. The model's ability to segment and track objects in complex environments surpassed conventional methods, making it particularly suitable for surveillance and security applications. The system consistently detected and tracked objects with exceptional precision, even in scenarios with occlusions, varying lighting conditions, and cluttered backgrounds. In this research work, we integrated the Segment Anything Model (SAM) with YOLO algorithm to present a novel method for training customised item identification models. When SAM and YOLO were used together, increased object detection and tracking abilities were shown, successfully overcoming issues such as complicated scenes, occlusions, and irregular object shapes. Through comprehensive testing, we confirmed the effectiveness of the suggested approach. We trained a customised YOLO model that attained high detection accuracy by downloading a sizable dataset of photos for a particular object category and using SAM for labelling. Instance segmentation from SAM and object recognition from YOLO were combined to produce reliable and accurate object localization. The suggested method has a number of benefits, including adaptation to different object types and scalability for busi- ness applications. The same procedure may be used to train customised models for other items using SAM, improving detection accuracy for each distinct object class. The method shows promise in areas including asset tracking, security surveillance, and theft avoidance. In summary, this study has significantly advanced the fields of object detection and tracking.[38] We have created a useful and effective method for training specific object identification models by combining SAM and YOLO. The outcomes show how the suggested methodology can be used to address item localization issues while attaining high detection accuracy.[33] Loss Evaluation: The graph that is displayed highlights the model's linear development. The training loss was dramatically decreased to just 0.05 by the 10th epoch, indicating the model's ability to assimilate the input. Training and validation losses line up, indicating that the model is not just learning well but also adjusting to new data with ease.

Figure 6: Progression of precision and F1 score across epochs chart

Analysis of Precision: The model's increasing accuracy and reliability are demonstrated by the steady increase in both precision and F1 score. The precision admirably increased to 0.98 by the tenth epoch, indicating a high percentage of accurate identifications. Simultaneously, a noteworthy 0.98 was obtained for the F1 score, which represents the harmonic average of precision and recall, underscoring the model's comprehensive effectiveness.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 7, Issue 11, pp 160-169, November-2023

https://doi.org/10.47001/IRJIET/2023.711023

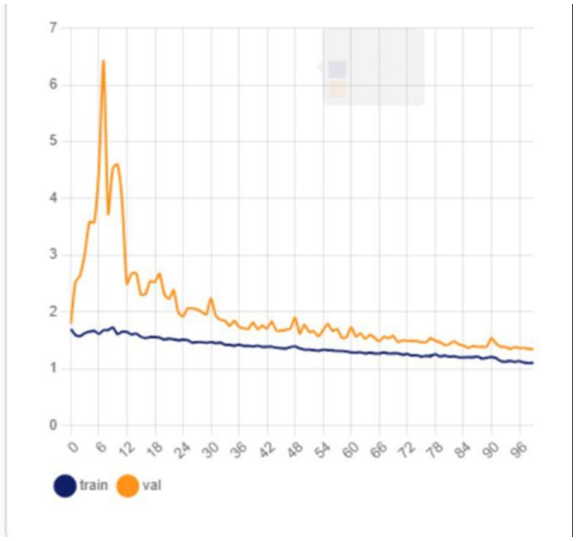

Figure 7: Object detection model accuracy

In conclusion, our approach to object detection and tracking, coupled with theft prevention and pose detection using a combination of classifiers, has yielded promising results. This integrated system offers a robust solution for enhancing security, monitoring human activity, and ensuring safety in diverse contexts, with potential applications ranging from retail and public spaces to healthcare and beyond.[45]

VI. CONCLUSION AND FUTURE PLANNING

A) Pose Detection and Theft Avoidance

The conclusion of the second article focuses on the construction of a complete security system that integrates sophisticated posture detection algorithms and theft prevention measures. The system solves security problems in institutional areas by properly recording human activity, identifying theft attempts, and providing real-time notifications to security staff. The research project proposes a fundamental solution that connects current technology with institutional security requirements. The combination of cutting-edge pose recognition algorithms, such as MediaPipe, with CNN models improves the system's capacity to properly detect human postures and motions.[35] The technology's proactive approach to theft prevention, considering both hand mobility and hand object contact, indicates its potential to drastically reduce security threats.

B) Integration of Object Detection and Tracking

The study results in the development of a unique strategy for developing customised item recognition models by combining the Segment Anything Model (SAM) with the YOLO algorithm. The combination of SAM's accurate labelling and YOLO's extensive recognition skills produced encouraging results. The technique demonstrated enhanced object recognition and tracking by overcoming hurdles given by intricate sceneries, occlusions, and irregular object forms. The method's versatility across several object categories, as well as its scalability for commercial applications, highlight its importance. The ability of SAM to adapt models for specific item classes, as well as its promise for asset monitoring, security surveillance, and theft prevention, opens the door to practical deployment.[39]

1) Future Planning: While the approach of each research indicates success, both conversations admit potential for improvement. Future research will focus on overcoming restrictions such as dataset variety and improving technique robustness in difficult settings. Overcoming these limitations can help to improve the techniques' robustness and applicability. Finally, the talks highlight the revolutionary potential of combining sophisticated algorithms for greater institutional security. Whether in the context of item detection and tracking or advanced algorithms in posture recognition, the research helps to create safer settings by leveraging technology's ability to precisely identify, track, and prevent thefts and suspicious behaviour.

REFERENCES

[1] A Review of Yolo Algorithm Developments. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050922001363 [2]Bo Wan, Desen Zhou, Yongfei Liu, Rongjie Li, Xuming He; Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 9469-9478.

[2] Wang, Limin, et al. "Temporal segment networks: Towards good practices for deep action recognition." Proceedings of the European Conference on Computer Vision. 2018.

[3] R. Punmiya and S. Choe, "Energy Theft Detection Using Gradient Boosting Theft Detector With Feature Engineering-Based Preprocess- ing," in IEEE Transactions on Smart Grid, vol. 10, no. 2, pp. 2326-2329, March 2019, doi: 10.1109/TSG.2019.2892595.

[4] M. Kisantal, S. Sharma, T. H. Park, D. Izzo, M. M a¨rtens and S. D'Amico, "Satellite Pose Estimation Challenge: Dataset, Competi- tion Design, and Results," in IEEE Transactions on Aerospace and Electronic Systems, vol. 56, no. 5, pp. 4083-4098, Oct. 2020, doi: 10.1109/TAES.2020.2989063.

[5] G. Rogez, P. Weinzaepfel and C. Schmid, "LCR-Net++: Multi-Person 2D and 3D Pose Detection in Natural Images," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 42, no. 5, pp. 1146- 1161, 1 May 2020, doi: 10.1109/TPAMI.2019.2892985.

ISSN (online): 2581-3048
Volume 7, Issue 11, pp 160-169, November-2023
https://doi.org/10.47001/IRJIET/2023.711023

[6] Pauzi, A.S.B. et al. (2021). Movement Estimation Using Mediapipe Blaze Pose. In: Badioze Zaman, H., et al. Advances in Visual Informatics. IVIC 2021. Lecture Notes in Computer Science(), vol 13051. Springer, Cham. https://doi.org/10.1007/978-3-030-90235-3-49.

[7] X. Yang, H. Sun, X. Sun, M. Yan, Z. Guo and K. Fu, "Position Detection and Direction Prediction for Arbitrary-Oriented Ships via Multitask Rotation Region Convolutional Neural Network," in IEEE Access, vol. 6, pp. 50839-50849, 2018, doi: 10.1109/ACCESS.2018.2869884.

[8] Yin Li, Xiaodi Hou, Christof Koch, James M. Rehg, and Alan L Yuille. 2014. . The secrets of salient object segmentation. In Proceedings of the IEEE Conference.

[9] A Comparative Study of Multiple Object Detection Using Haar-Like Feature Selection and Local Binary Patterns in Several Platforms. [On- line]. Available: https://www.hindawi.com/journals/mse/2015/948960/

[10] Real-Time Tracking of Multiple People Using Continuous Detection. [Online].

[11] R. De Geest and T. Tuytelaars, "Modeling Temporal Structure with LSTM for Online Action Detection," 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, NV, USA, 2018, pp. 1549-1557, doi: 10.1109/WACV.2018.00173.

[12] C. Peng and Q. Cheng, "Discriminative Ridge Machine: A Classifier for High-Dimensional Data or Imbalanced Data," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 6, pp. 2595-2609, June 2021, doi: 10.1109/TNNLS.2020.3006877.

[13] A survey of techniques for human detection from video. [Online]. Avail- able: http://www.cs.umd.edu/grad/scholarlypapers/papers/neetiPaper.pdf

[14] J. Yun et al., "Grasping Pose Detection for Loose Stacked Ob- ject Based on Convolutional Neural Network with Multiple Self- Powered Sensors Information," in IEEE Sensors Journal, 2022, doi: 10.1109/JSEN.2022.3190560.

[15] YOLO-ACN: Focusing on Small Target and Occluded Object Detection. [Online].

[16] Hetang, Congrui, et al. "Impression network for video object detection." Proceedings of the European Conference on Computer Vision. 2020. [18]Chuan Yang, Lihe Zhang, Huchuan Lu, Xiang Ruan, and Ming-Hsuan Yang. 2013. Saliency detection via graph-based.

[17] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. [Online]. Available: https://arxiv.org/pdf/1712.05896.pdf

[18] C. Lugaresi, J. Tang, H. Nash, C. McClanahan, E. Uboweja, M. Hays, F. Zhang, C.-L. Chang, M. Yong, J. Lee, W.-T. Chang, W. Hua, M. Georg, M. Grundmann (2019) MediaPipe: a framework for building perception pipelines.

[19] Feichtenhofer, Christoph, et al., "Detect to track and track to detect." Proceedings of the IEEE International Conference on Computer Vision. 2017.

[20] Sun, Xingyi, et al. "Deep high-resolution representation learning for visual recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019.

[21] F. Zhang, V. Bazarevsky, A. Vakunov, A. Tkachenka, G. Sung, C.L. Chang, and M. Grundmann, "MediaPipe Hands: On-device Real- time Hand Tracking," in Proceedings of CVPR Workshop on Computer Vision for Augmented and Virtual Reality, Seattle, WA, USA, 2020, pp. 5 pages, 7 figures. [Online]. Available: https://arxiv.org/abs/2006.10214

[22] Human Detection Using Oriented Histograms of Flow and Appearance. [Online].

[23] W. Shi, J. Cao, Q. Zhang, Y. Li and L. Xu, "Edge Computing: Vision and Challenges," in IEEE, 2020.

[24] T. Baltru sˇaitis, P. Robinson and L.-P. Morency, "OpenFace: An open source facial behavior analysis toolkit," in IEEE Transactions on Bio- metrics, Behavior, and Identity Science, 2022.

[25] F. D. P.,. Z. C. L. Chaoning Zhang, "A Survey on Segment Anything Model (SAM): Vision Foundation Model Meets Prompt Engineering," 2023.

[26] H. N.-p. 1. R. a. R. D. Behzad Mirzaei, "Small Object Detection and Tracking: A Comprehensive Review," 2023.

[27] W. Li1, "Analysis of Object Detection Performance Based on Faster RCNN," in Journal of Physics: Conference Series, 2021.

[28] C. Wan, "Gait Recognition Based on Deep Learning: A Survey," Brazil, 2022.

[29] M. Mehra, V. Sahai, P. Chowdhury and E. Dsouza, "Home Security System using IOT and AWS Cloud Services," in IEEE Trust- com/BigDataSE/ICESS, 2017.

[30] "Local ridge regression for face recognition," in Hui Xue, Yulian Zhu, 2009.

[31] D. Sudiana, M. Rizkinia and F. Alamsyah, "Performance Evaluation of Machine Learning Classifiers for Face Recognition," in IEEE, Depok, Indonesia, 2021.

ISSN (online): 2581-3048
Volume 7, Issue 11, pp 160-169, November-2023
https://doi.org/10.47001/IRJIET/2023.711023

[32] R. O. Ogundokun, R. Maskeli u¯nas, and R. Damas˘evic˘ius, "Human Posture Detection Using Image Augmentation and Hyperparameter- Optimized Transfer Learning Algorithms," Applied Sciences, vol. 12, no. 19, p. 10156, Oct. 2022, doi: https://doi.org/10.3390/app121910156.

[33] A.K. Raja, C. Sugandhi, G. Nymish, N. S. Havish and M. Rashmi, "Face Gesture Based Virtual Mouse Using Mediapipe," in IEEE Xplore, Lonavla, India, 2023.

[34] R. Girshick, J. Donahue, T. Darrell and J. Malik, "Rich Feature Hier- archies for Accurate Object Detection and Semantic Segmentation," in IEEE, Columbus, OH, USA, 2014.

[35] S. D. R. G. A. F. Joseph Redmon, "You Only Look Once: Unified, Real-Time Object Detection," 2015.

[36] O. J. M. S. Authors: Alper Yilmaz, "Object tracking: A survey," in Association for Computing Machinery, New York, NY, United States, 2006.

[37] K. H. a. Y. G. Ujwalla Gawande, "Pedestrian Detection and Tracking in Video Surveillance System: Issues, Comprehensive Review, and Challenges," in intechopen, 2019.

[38] L. E. C. a. M. Felson, "Social Change and Crime Rate Trends: A Routine Activity Approach," in American Sociological Review.

[39] C. Di Natali, M. Beccani and P. Valdastri, "Real-Time Pose Detection for Magnetic Medical Devices," in IEEE Transactions on Magnetics, vol. 49, no. 7, pp. 3524-3527, July 2013, doi: 10.1109/TMAG.2013.2240899.

[40] S. Das, Md. S. Imtiaz, N. H. Neom, N. Siddique, and H. Wang, "A hybrid approach for Bangla sign language recognition using deep transfer learning model with random forest classifier," Expert Systems with Applications, p. 118914, Sep. 2022, doi: https://doi.org/10.1016/j.eswa.2022.118914.

[41] D. Sethi, S. Bharti, and C. Prakash, "A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work," Artificial Intelligence in Medicine, p. 102314, May 2022, doi: https://doi.org/10.1016/j.artmed.2022.102314.

[42] A.F. Abate, C. Bisogni, A. Castiglione, and M. Nappi, "Head pose estimation: An extensive survey on recent techniques and ap- plications," Pattern Recognition, vol. 127, p. 108591, Jul. 2022, doi: https://doi.org/10.1016/j.patcog.2022.108591.

[43] S. Garg, A. Saxena, and R. Gupta, "Yoga pose classification: a CNN and MediaPipe inspired deep learning approach for real-world application," Journal of Ambient Intelligence and Humanized Computing, Jun. 2022, doi: https://doi.org/10.1007/s12652-022-03910-0.

[44] M.Park, D. Q. Tran, J. Bak, and S. Park, "Small and overlapping worker detection at construction sites," Automation in Construction, vol. 151, p. 104856, Jul. 2023, doi: https://doi.org/10.1016/j.autcon.2023.104856.

Citation of this Article:

Chathuranga K. G. S, Vidanage K. H, Dr. Harinda Fernando, Dr. Lakmini Abeywardhana, "Human, Object and Pose Detection for Theft Prevention through Surveillance System" Published in *International Research Journal of Innovations in Engineering and Technology - IRJIET*, Volume 7, Issue 11, pp 160-169, November 2023. Article DOI https://doi.org/10.47001/IRJIET/2023.711023

****

