A

Project Report

on

Human Detection to Mitigate Excessive Electrical Energy Usage in Communal Spaces

submitted as partial fulfillment for the award of

BACHELOR OF TECHNOLOGY DEGREE

SESSION 2023-24

in

Computer Science and Engineering

By

Aabhas Bisaria (2000290100001)

Devesh Chauhan (2000290100054)

Harsh Varshney (2000290100065)

Under the supervision of

Dr. Sushil Kumar

KIET Group of Institutions, Ghaziabad

Affiliated to

Dr. A.P.J. Abdul Kalam Technical University, Lucknow

(Formerly UPTU)

May,2024

TABLE OF CONTENTS

DECLARATION

We hereby declare that this submission is our own work and that, to the best of our knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning, except where due acknowledgment has been made in the text.

Signature:

Name:

Roll No.:

Date:

Signature:

Name:

Roll No.:

Date:

Signature:

Name:

Roll No.:

Date:

CERTIFICATE

This is to certify that Project Report entitled “Human Detection to Mitigate Excessive Electrical Energy Usage in Communal Spaces” which is submitted by Aabhas Bisaria, Devesh Chauhan and Harsh Varshney in partial fulfillment of the requirement for the award of degree B.Tech. in Department of Computer Science & Engineering of Dr. A.P.J. Abdul Kalam Technical University, Lucknow is a record of the candidates own work carried out by them under my supervision. The matter embodied in this report is original and has not been submitted for the award of any other degree.

.

Dr. Sushil Kumar	**Dr. Vineet Sharma**
Associate Professor	**(Head of Department)**

Date:

ACKNOWLEDGEMENT

It gives us a great sense of pleasure to present the report of the B. Tech Project undertaken during B. Tech. Final Year. We owe special debt of gratitude to Dr. Sushil Kumar, Department of Computer Science & Engineering, KIET, Ghaziabad, for his constant support and guidance throughout the course of our work. His sincerity, thoroughness and perseverance have been a constant source of inspiration for us. It is only his cognizant efforts that our endeavors have seen light of the day.

We also take the opportunity to acknowledge the contribution of Dr. Vineet Sharma, Head of the Department of Computer Science & Engineering, KIET, Ghaziabad, for his full support and assistance during the development of the project. We also do not like to miss the opportunity to acknowledge the contribution of all the faculty members of the department for their kind assistance and cooperation during the development of our project.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members, especially Dr. Sushil Kumar, of the CSE department for their kind assistance and cooperation during the development of our project. Last but not the least, we acknowledge our friends for their contribution in the completion of the project.

Date:

Signature:

Name:

Roll No.:

Date:

Signature:

Name:

Roll No.:

Date:

Signature:

Name:

Roll No.:

ABSTRACT

The study aims to delve deeply into the escalating issue of electrical energy consumption within communal settings, particularly focusing on high-traffic areas such as restrooms and multiplexes. These locations serve as crucial hubs of human activity, but their energy consumption patterns often go unchecked, leading to significant financial burdens for both users and facility operators. Moreover, the unchecked energy usage exacerbates strain on already burdened energy grids, resulting in heightened carbon emissions and exacerbating environmental concerns such as climate change.

To address this pressing issue, the study proposes a novel approach leveraging the TensorFlow Object Counting API. This innovative strategy seeks to develop a sophisticated system capable of accurately and efficiently counting human subjects in real-time within these communal spaces. By precisely identifying and tallying the number of occupants present, the system aims to curtail unnecessary electricity consumption, thereby mitigating the financial and environmental impacts associated with excessive energy usage.

The proposed system is underpinned by the robust capabilities of the TensorFlow platform, offering a versatile and user-friendly framework for developing advanced object counting systems. Through intricate algorithms and state-of-the-art technology, the system demonstrates remarkable proficiency in item identification, tracking, and counting, showcasing its adaptability and resilience in addressing complex challenges.

By tackling the dual challenges of rising electricity bills and environmental degradation, the proposed technology presents a promising solution to the problem at hand. Through meticulous research and analysis, the study seeks to elucidate the underlying factors contributing to excessive energy consumption in communal settings. Furthermore, it aims to propose practical and actionable solutions to effectively manage and mitigate energy wastage.

Ultimately, the overarching goal of the study is to implement measures that promote sustainable resource management and environmental conservation in communal spaces. By optimizing energy usage and fostering a culture of efficiency, the study endeavors to contribute to the collective efforts aimed at building a more sustainable and environmentally conscious society.

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ML	Machine Learning
CNN	Convolutional Neural Networks
ANN	Artificial Neural Networks
SSD	Single Shot Detector
R-CNN	Regional - Based Convolutional Neural Networks
FNP	Featured Pyramid Networks
YOLO	You Only Look Once
API	Application Programming Interface
NumPy	Numerical Python
GPU	Graphics Processing Unit
CPU	Central Processing Unit
RAM	Random Access Memory
Open CV	Open Source Computer Vision Library

CHAPTER 1

INTRODUCTION

1.1 Introduction

In the current setting, one of the most important criteria that must be satisfied is to make the most of the available resources and to spend them as efficiently as possible. Currently, the problem at hand is the unregulated and ongoing consumption of electrical energy , which frequently surpasses the requirements that are being met. Because it raises a considerable concern, this constitutes a significant problem that calls for an immediate solution. There is a wide variety of various communal settings in which the observed predominance of wastefulness is apparent. Restrooms, public restrooms, and multiplexes are some examples of these types of locations; however, this list is not exhaustive. The analysis of the data reveals a discernible pattern in which the lighting fixtures located within the designated areas consistently tend to remain activated, even when there are no occupants present. This is particularly obvious in situations where there are no people around. Even when there are no other individuals in the vicinity, this is still the case. To solve the issue of increased energy consumption, the primary objective of the current research is to offer a real-time human counter system as a means of tackling the problem. The current system has been painstakingly constructed with the idea of accomplishing the primary objective of optimizing energy efficiency through the removal of superfluous electricity usage. This target has been the fundamental inspiration for the construction of the system. When it comes to creativity and originality, the open-source framework that is currently being evaluated demonstrates an impressive amount of exceptional qualities. To construct this framework, the TensorFlow platform was utilized. The system provides users with a methodology that is not only user-friendly but also very efficient. This is done to simplify the creation of sophisticated object-counting systems. The system that is the focus of this investigation has an outstanding level of adaptability and efficiency in its operation. The resilient capabilities of the system in the areas of object recognition, tracking, and counting may be principally responsible for this particular consequence. Taking into consideration the possibilities of this technology, there is a significant likelihood that it will bring about a revolutionary change in the management and operation of lighting systems in public spaces. The technology that is being investigated demonstrates promising capabilities in properly recognizing the presence of individuals in public places and accurately monitoring the movements of those individuals. It is being taken into mind that certain skills are available. In addition to this, it will ensure that the lighting conditions within these specific regions are not only adequate but also excellent in terms of the amount of energy that they consume. When it comes to the study that we are conducting, our major purpose is not simply to cut down on the amount of electricity that is used. The methodology that has been shown may be a unique approach that has the potential to result in considerable reductions in both the amount of energy that is used and the costs that are associated with both of these things. A thorough analysis of our real-time human counting system is going to be carried out in the subsequent sections, which will thereafter be presented. It will be important to perform a comprehensive examination of the technological components to properly execute the endeavor that has been recommended. This investigation will encompass the intricate mechanisms that are involved in the process of object recognition and tracking. As an additional point of interest, the primary objective of this research endeavor is to conduct a comprehensive examination into the combination of the system that was mentioned earlier with various techniques for the management of lighting. A thorough examination will also be carried out to evaluate the possible implications of this technology concerning the conservation of electricity in a range of public contexts. This investigation will be carried out to examine the potential implications of this technology. The fundamental objective of the current study, which intends to provide in-depth information on these implications, is to provide a thorough understanding of the larger implications of our research in effectively addressing the existing challenges related to energy conservation. This is the primary objective of the study.

1.2 Project Description

The project embarks on a vital mission to tackle the rampant issue of unregulated electrical energy consumption prevalent in numerous communal settings, where lighting fixtures persistently remain active even in unoccupied spaces. This wasteful pattern of energy usage not only exceeds actual demand but also poses significant concerns regarding sustainability and resource optimization. Recognizing the urgency of this challenge, the research endeavors to introduce a real-time human counter system as a proactive measure to optimize energy efficiency and curb unnecessary electricity usage. At its core, the proposed solution is driven by a commitment to maximizing resource utilization and spending available resources as efficiently as possible. By leveraging cutting-edge technology and innovative methodologies, the project aims to revolutionize the management of lighting systems in public spaces. Central to this endeavor is the development of a sophisticated framework built upon the robust foundations of TensorFlow, an open-source platform renowned for its versatility and efficiency in constructing advanced object-counting systems. This framework not only promises exceptional adaptability but also demonstrates remarkable efficiency in object recognition, tracking, and counting.

Through meticulous analysis and experimentation, the research seeks to harness the full potential of this technology to accurately discern human presence and monitor movements in public areas. By integrating the real-time human counter system with various lighting management techniques, the project endeavors to ensure lighting conditions are not only adequate but also optimized in terms of energy consumption. This holistic approach not only addresses the immediate need to reduce electricity usage but also lays the groundwork for sustainable energy practices in the long term.

Beyond mere energy reduction, the project aspires to pioneer a unique approach capable of yielding substantial reductions in both energy consumption and associated costs. By conducting a comprehensive examination of the technological components and implications of the proposed solution, the research aims to provide invaluable insights into its transformative potential in mitigating energy conservation challenges. Furthermore, the project is committed to fostering a deeper understanding of the broader implications of the research, particularly in effectively addressing existing challenges related to energy conservation in diverse public contexts.

In essence, the project represents a concerted effort to harness the power of technology and innovation to confront one of the most pressing challenges of our time. By developing and implementing a real-time human counter system, the research aims to not only optimize energy efficiency but also pave the way for more sustainable energy practices in public spaces. Through rigorous analysis and experimentation, the project endeavors to provide practical solutions that align with the overarching goal of resource optimization and efficiency.

CHAPTER 2

LITERATURE REVIEW

The field of computer vision has witnessed a remarkable evolution, with significant advancements in object detection methodologies over the past decade. Among these advancements, the Single Shot Detector (SSD) emerged as a groundbreaking approach, departing from traditional two-shot detectors like R-CNN. By unifying regional proposal generation and object detection, SSD introduced a paradigm shift, particularly in scenarios prioritizing convenience and efficiency. Similarly, the introduction of the You Only Look Once (YOLO) algorithm presented a unified model that revolutionized real-time object detection. This model, by analyzing the entire image in a single pass, eliminated the need for separate object location and classification, resulting in faster and more efficient detection.

Another milestone in computer vision was marked by the introduction of Mask R-CNN, which demonstrated significant progress in instance segmentation. Building upon the Faster R-CNN framework, Mask R-CNN pioneered a novel architecture capable of predicting object bounding boxes, class labels, and pixel-level masks simultaneously. This approach facilitated precise object segmentation at the pixel level, a crucial advancement for applications requiring accurate object placement and separation. The integration of pixel-level mask prediction into an object detection framework represented a departure from traditional methods, establishing Mask R-CNN as an industry standard for pixel-by-pixel segmentation.

Girshick's Fast R-CNN further contributed to the field by introducing a unified framework that merged region proposal generation and object identification into a single model. This model addressed computational inefficiencies present in its predecessors and significantly impacted the field of object detection, inspiring subsequent research and becoming a reference point for new algorithms. The versatility of Fast R-CNN in handling region proposal and object detection tasks has found applications in various domains beyond academia, including surveillance, automatic systems, and image-based searching applications.

Addressing the challenge of scale variance in object recognition, Feature Pyramid Networks (FPN) introduced a novel framework for capturing features at multiple scales. By forming a feature pyramid and leveraging convolutional networks, FPN enables accurate and scalable object recognition by pairing high-level semantic properties with precise spatial information. This architectural innovation has contributed significantly to the field, enhancing the accuracy and scalability of object detection systems.

2.1 Advantage of Single Shot Detector (SSD)

The introduction of the Single Shot Detector (SSD) marks a pivotal advancement in the realm of object detection within computer vision. Traditionally, object detection systems relied on a two-shot approach, exemplified by methods like region-based Convolutional Neural Networks (R-CNN), where separate stages were employed for region proposal and subsequent object classification. However, SSD revolutionized this paradigm by amalgamating these processes into a singular model, thereby streamlining, and accelerating the entire detection pipeline.

At the core of SSD's innovation lies its capability to generate region proposals and conduct object detection concurrently in a single forward pass of the neural network. Unlike its predecessors, SSD eliminates the need for external region proposal algorithms, rendering it more efficient and expeditious. This integration of regional proposal generation and object detection not only enhances efficiency but also significantly accelerates the detection process, enabling real-time performance even on devices with limited computational resources.

The significance of SSD extends far beyond its technical prowess; it represents a paradigm shift in object detection methodologies. By consolidating multiple stages into a unified framework, SSD achieves a delicate balance between accuracy and speed, making it particularly well-suited for applications where real-time detection is paramount. Industries such as autonomous driving, video surveillance, and augmented reality stand to benefit immensely from SSD's capabilities, as it enables swift and accurate identification of objects in dynamic environments.

Moreover, SSD's versatility and effectiveness have garnered widespread adoption within the computer vision community, driving further research and development in the field of real-time object detection. Its success has inspired a wave of innovations aimed at pushing the boundaries of detection speed and accuracy, paving the way for enhanced applications across various domains.

In summary, the advent of SSD represents a monumental leap forward in object detection technology. By unifying region proposal generation and object detection into a single model, SSD not only enhances efficiency but also unlocks new possibilities for real-time applications in diverse fields. Its impact reverberates throughout the computer vision community, shaping the trajectory of research and development in pursuit of more efficient and effective object detection solutions.

2.2 Unifying Region Proposal and Object Identification: Fast R-CNN

Traditionally, object detection systems employed a two-stage approach. First, they generated regional proposals to identify potential object locations within an image. Then, these proposals were passed through a separate classifier to determine the presence and category of objects. However, this approach was computationally expensive and inefficient, often resulting in slow inference times.

Fast R-CNN addressed these limitations by introducing a unified framework that streamlined the entire detection pipeline. Instead of relying on external algorithms for region proposal generation, Fast R-CNN integrated a Region of Interest (RoI) pooling layer directly into the network architecture. This layer allowed the model to extract fixed-size feature maps from arbitrary-sized regions of the input image, enabling it to focus on relevant regions without the need for additional computations.

By unifying region proposal generation and object identification, Fast R-CNN achieved significant improvements in both speed and accuracy. The model could efficiently process entire images in a single forward pass, eliminating the need for redundant computations and dramatically reducing inference times. Additionally, by sharing convolutional features across region proposals, Fast R-CNN improved feature reuse and representation learning, leading to more robust and discriminative object detections.

The impact of Fast R-CNN on the field of object detection cannot be overstated. Its introduction marked a paradigm shift in detection methodologies, inspiring a wave of research and development in unified detection frameworks. Furthermore, Fast R-CNN served as a reference point for subsequent algorithms, including Faster R-CNN, which further refined the unified detection approach by introducing a separate Region Proposal Network (RPN) to improve proposal generation efficiency.

Beyond academia, Fast R-CNN found widespread application in various domains, including surveillance, automatic systems, and image-based searching applications. Its efficiency and effectiveness made it a go-to choice for real-world deployment, driving further adoption and innovation in the field of object detection.

2.3 Architectural Innovation: Building Feature Pyramids for Object Recognition

Feature Pyramid Networks (FPN) address a fundamental challenge in object recognition known as scale variance. Traditional approaches to object detection struggle with accurately detecting objects at different scales within an image. Objects may appear in various sizes and resolutions, making it challenging for the model to detect and classify them consistently across the image.

FPN tackles this challenge by introducing a novel architecture that captures features at multiple scales simultaneously. At its core, FPN constructs a feature pyramid by leveraging a single convolutional network. This pyramid consists of multiple feature maps, each capturing information at a specific scale, ranging from low to high resolutions. These feature maps serve as building blocks for the pyramid, with each level containing increasingly detailed information about the image.

The design of FPN enables high-level semantic properties to be paired with precise spatial information, allowing for more accurate and scalable object recognition. By incorporating features from multiple scales, FPN overcomes the limitations of traditional approaches and provides a comprehensive understanding of the image's content across different resolutions.

One of the key contributions of FPN is its ability to perform instance segmentation, a task that requires precise delineation of object boundaries at the pixel level. By extending the Faster R-CNN framework with a parallel branch for segmentation mask prediction, FPN enables precise object segmentation, which is crucial for applications requiring accurate object placement and separation.

Moreover, FPN represents a departure from traditional approaches by integrating pixel-level mask prediction into the object detection framework. This architectural innovation enhances the model's versatility and effectiveness, paving the way for more accurate and scalable object detection systems.

The impact of FPN extends beyond its technical innovations. Its introduction has catalyzed advancements in object recognition and instance segmentation, driving further research and development in the field. FPN has become a cornerstone in the arsenal of tools used by researchers and practitioners in computer vision, shaping the trajectory of object detection methodologies.

2.4 Advancing Precision: Mask R-CNN and Instance Segmentation

Mask R-CNN stands as a groundbreaking advancement in the realm of computer vision, particularly in the field of object detection and instance segmentation. Traditionally, object detection systems focused on identifying objects within an image and predicting their bounding boxes, serving critical roles in applications like object recognition, scene understanding, and autonomous navigation. However, while these systems excelled in identifying objects at a coarse level, they often struggled with accurately delineating object boundaries, especially in scenarios where precise segmentation was required.

Recognizing this limitation, researchers endeavored to develop more robust methodologies capable of not only detecting objects but also precisely segmenting them at the pixel level. This led to the inception of Mask R-CNN, an innovative architecture that builds upon the success of the Faster R-CNN framework while introducing novel components to enable pixel-level instance segmentation.

At the core of Mask R-CNN's architecture lies its ability to predict three key components simultaneously: object bounding boxes, class labels, and pixel-level segmentation masks. Unlike traditional object detection systems, which relied solely on bounding box predictions, Mask R-CNN incorporates a parallel branch dedicated to pixel-wise mask prediction. This branch operates in conjunction with the existing branches for class classification and bounding box regression, allowing the model to produce accurate segmentation masks for each detected object instance.

The introduction of pixel-level segmentation masks represents a significant breakthrough in object detection methodologies, enabling Mask R-CNN to precisely delineate object boundaries at the finest granularity. This capability is particularly invaluable in scenarios where objects overlap or exhibit intricate shapes and structures, such as in medical imaging, industrial inspection, and scene understanding for robotics applications.

Furthermore, the versatility of Mask R-CNN extends beyond its technical innovations, impacting a wide range of applications and domains. In fields like image editing and augmented reality, Mask R-CNN's instance segmentation capabilities enable advanced functionalities like object manipulation, virtual object insertion, and background replacement with unparalleled precision. Similarly, in medical imaging, Mask R-CNN facilitates the segmentation of anatomical structures and lesions with high accuracy, aiding in disease diagnosis and treatment planning.

The success of Mask R-CNN has spurred further research and development in the field of instance segmentation, driving advancements in algorithmic efficiency, accuracy, and scalability. Its impact reverberates across academia and industry, shaping the trajectory of computer vision research and paving the way for more sophisticated and versatile object detection systems.

2.5 Revolutionizing Real-Time Object Detection with YOLO Algorithm

The introduction of the You Only Look Once (YOLO) algorithm has marked a significant breakthrough in the domain of object detection within computer vision. This innovative approach has redefined the landscape of real-time object detection by introducing a unified framework that analyzes the entire image in a single pass, revolutionizing traditional methodologies that relied on multi-stage processing.

In conventional object detection systems, images were processed through multiple stages, involving intricate pipelines for region proposal and subsequent object classification. These multi-pass approaches often led to computational inefficiencies and slower inference times, particularly in applications requiring real-time performance. YOLO, however, challenged this paradigm by streamlining the detection process into a single, efficient pass through the network.

The architecture of YOLO is characterized by its simplicity and effectiveness. By dividing the input image into a grid and making predictions for each grid cell simultaneously, YOLO eliminates the need for complex region proposal algorithms. This unified approach not only reduces computational overhead but also enables real-time performance, even on devices with limited computational resources.

One of the key advantages of YOLO is its ability to achieve real-time object detection without compromising on accuracy. By analyzing the entire image at once, YOLO ensures that objects are detected promptly and with high precision. This capability makes YOLO particularly well-suited for applications where timely and accurate detection of objects is critical, such as surveillance systems, autonomous vehicles, and augmented reality.

Furthermore, the versatility of YOLO extends across a wide range of applications and domains. Its real-time object detection capabilities find applications in diverse fields, including video analysis, real-time tracking, and object recognition in dynamic environments. YOLO's efficiency and accuracy make it a valuable tool for researchers and practitioners alike, driving advancements in computer vision research and applications.

The success of YOLO has spurred further research and development in the field of real-time object detection. Its innovative architecture and impressive performance have inspired a wave of innovations in unified detection frameworks, leading to advancements in computer vision research and applications. The impact of YOLO transcends academia, shaping the landscape of object detection technologies and paving the way for more efficient and effective detection systems.

CHAPTER 3

PROPOSED METHODOLOGY

The proposed methodology is meticulously designed to address the critical need for accurately quantifying human movement within specific locations. At its core, the system focuses on two pivotal functionalities: quantifying the influx and efflux of individuals entering and exiting a designated area. Achieving this objective requires a tailored approach, which involves the utilization of both a human detection module and a tracking mechanism.

Human detection, a cornerstone of the methodology, relies on the deployment of bounding boxes. These bounding boxes serve as visual representations, encapsulating recognized human entities within the scene. Each bounding box is endowed with a unique identification number, crucial for tracking and monitoring purposes. By assigning distinct IDs to individual objects, the system can effectively monitor their trajectories and spatial displacements over time.

Central to the proposed methodology is the integration of cutting-edge computer vision techniques. Specifically, the system harnesses the power of the Single Shot Detector (SSD) in conjunction with the Mobile Net architecture for real-time object detection. This strategic amalgamation optimizes both speed and efficiency, enabling rapid and precise detection of objects, including humans. Unlike conventional methods, SSD operates with remarkable efficiency, capable of identifying objects in a single iteration. This streamlined approach eliminates the need for sequential region proposal and object identification processes, resulting in enhanced performance and speed.

Furthermore, the methodology incorporates centroid tracking as a fundamental component. Centroid tracking enables the system to monitor objects across multiple frames by determining the centroid of bounding boxes generated by the SSD. This technique facilitates accurate tracking of object trajectories and spatial coordinates, ensuring reliable monitoring of human movement within the designated area.

In practical terms, each detected object is endowed with a unique identification (ID), facilitating seamless tracking, and recording of object trajectories. This continuous monitoring process ensures consistency and enables the system to reconstruct the movement patterns of individual objects over time accurately.

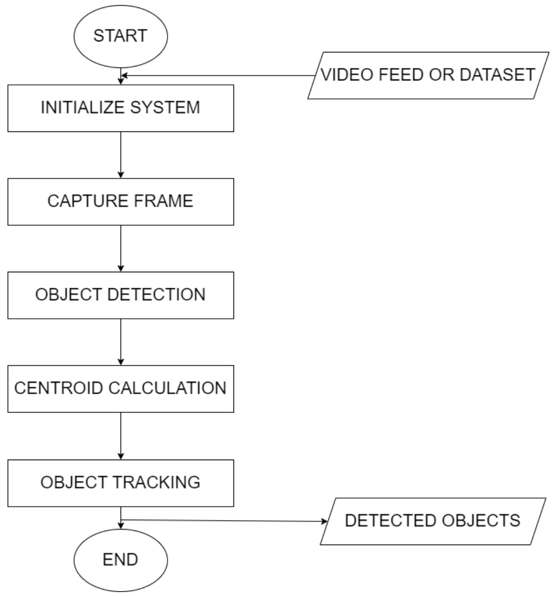


Fig. 1. Flowchart of proposed work

3.1 Human Detection and Tracking Functionality

The proposed system focuses on accurately quantifying the number of individuals entering and exiting a specific location through robust human detection and efficient tracking mechanisms.

The human detection module, equipped with sophisticated algorithms and machine learning techniques, analyzes visual data to identify human features amidst varying environmental conditions. By distinguishing human entities from other objects or background elements, the module ensures precise detection of individuals entering or traversing through the monitored area.

Complementing the human detection module is a sophisticated tracking mechanism that monitors the movement of recognized human objects in real-time. Once an individual is detected, the tracking mechanism continuously monitors their spatial trajectory and movement patterns within the monitored area, ensuring accurate and uninterrupted monitoring.

By combining the capabilities of the human detection module and the tracking mechanism, the system achieves comprehensive coverage of human activity within the monitored location, providing real-time insights into movement behaviors and patterns. This approach facilitates efficient crowd management, enhances security measures, and enables informed decision-making in various applications.

Overall, the system's emphasis on human detection and tracking underscores its commitment to accurately quantifying human presence and movement within specified locations, enabling effective resource management and optimization in diverse operational contexts.

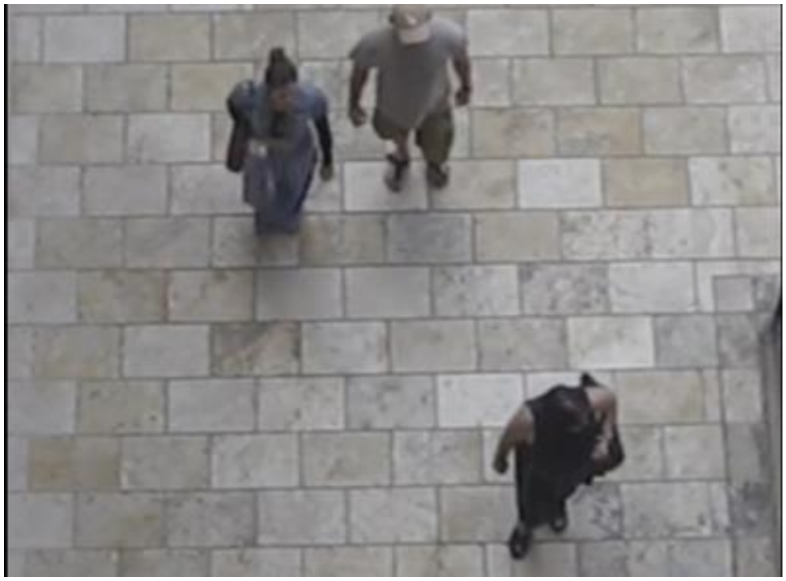


Fig. 2. Stock video frame without centroid allocation

3.2 Utilizing Bounding Boxes for Visual Representation

In the proposed system, human detection is facilitated through the utilization of bounding boxes, which serve as visual representations of recognized human entities within the monitored scene. These bounding boxes are essentially rectangular frames that encapsulate the detected individuals, providing a clear delineation of their presence in the environment.

Each bounding box is assigned a unique identification number, which serves as a reference point for tracking and monitoring purposes. This unique ID enables the system to differentiate between different individuals and track their movements accurately over time. By associating each bounding box with a distinct identifier, the system can maintain continuous surveillance of multiple individuals simultaneously, ensuring precise monitoring and analysis of human activity within the monitored area.

Overall, the utilization of bounding boxes for visual representation streamlines the process of human detection and tracking, providing a structured and efficient framework for monitoring human presence and movement in real-time.

Fig. 3. Stock video frame after centroid allocation

3.3 Utilizing Single Shot Detector (SSD) and Mobile Net Architecture

The methodology of the proposed system leverages computer vision techniques, specifically the Single Shot Detector (SSD) in conjunction with the Mobile Net architecture, to achieve real-time object detection.

SSD is a state-of-the-art object detection algorithm that can identify objects in a single pass through the network, eliminating the need for multiple stages of processing. This efficiency significantly speeds up the detection process compared to traditional methods, such as two-shot detectors like R-CNN, which require separate region proposal and object identification stages.

The Mobile Net architecture is a lightweight convolutional neural network (CNN) designed for efficient deployment on mobile and embedded devices. By integrating SSD with Mobile Net, the system optimizes both speed and efficiency, enabling rapid and accurate detection of objects, including humans, in real-time.

Overall, this methodology allows the system to efficiently process visual data, making it well-suited for applications requiring fast and accurate object detection, such as surveillance, autonomous driving, and augmented reality.

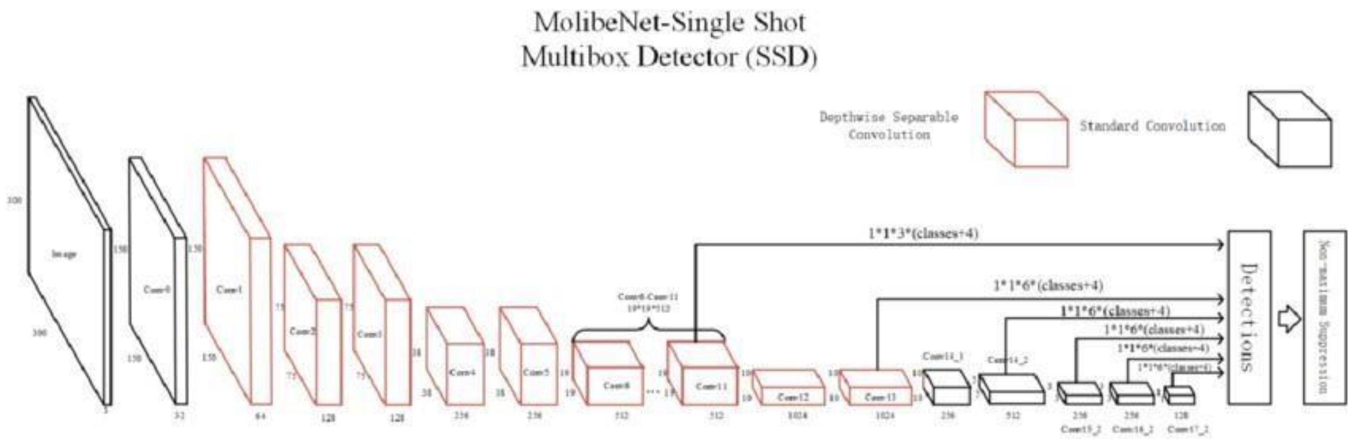


Fig. 4. Single Shot Detector (SSD) working

CHAPTER 4

RESULTS AND DISCUSSION

To accomplish the goals of this investigation, the system that is being studied is put through a series of tests on the graphics processing unit (GPU) of a computer that is located in the investigation area. The dataset that is being utilized in these experiments has been selected with considerable care. The central processing unit (CPU) was placed in the role of the target machine during the initial testing of the system. Regardless of this, it was quickly rejected because the GPU displayed substantially higher levels of performance than previously thought. Only the number of line incursions going up and down are significant for counting. Tables I and II display the accurate manual count and the predicted count using the proposed technique for various camera angles. To get the final accuracy percentage, calculate the frequency of error by subtracting the anticipated count from the observed count. For the more accurate normal model, the ultimate accuracy is 91.66%, while for the faster model, it is 78.33%. Both results are elaborately outlined and shown separately below.

Rate of error = sum of frequency of error/ (sum of anticipated increase and anticipated decrease) × 100%

= (0+1+1+1+2)/ (1+4+6+3+11+11+2+6+2+4)× 100%

= (5/50) × 100 %

= 0.1000 × 100%

= 10%

Overall accuracy = total anticipated count - rate of error

= 100% - 10%

= 90%

Accuracy outcome of standard model

Serial No.	Camera positioning	Anticipated increase	Anticipated decrease	Actual increase	Actual decrease
I	Anterior vision	1	4	1	4
II	Anterior vision	6	3	7	3
III	Top vision	11	11	11	12
IV	Top vision	2	6	2	5
V	Top vision	2	4	1	3

Rate of error = sum of frequency of error/ (sum of anticipated increase and anticipated decrease)× 100%

= (0+5+3+0+1)/ (1+4+6+3+11+11+2+6+2+4)× 100%

= (9/50) × 100%

= 0.18×100% = 18%

Overall accuracy = total anticipated count - rate of error

= 100% - 18%

= 82%

Accuracy outcome with increased speed

Serial No.	Camera positioning	Anticipated increase	Anticipated decrease	Actual increase	Actual decrease
I	Anterior vision	1	4	2	3
II	Anterior vision	6	3	3	1
III	Top vision	11	11	10	9
IV	Top vision	2	6	2	6
V	Top vision	2	4	2	5

CHAPTER 5

CONCLUSION AND FUTURE SCOPE

5.1 Conclusion

The research delves into a comprehensive exploration of computer vision, specifically focusing on the integration of the Single Shot Detector (SSD) approach with the Mobile Net architecture to enable real-time object detection. This combination presents a promising avenue for enhancing speed and efficiency, particularly in mobile devices, IP cameras, and scanners. Unlike conventional two-shot detectors like R-CNN, SSD boasts the capability to identify objects in a single iteration, merging region proposals and object identification into a unified phase. This strategic fusion eliminates the need for a separate region proposal network (RPN), thus streamlining processing, optimizing resources, and simplifying the detection pipeline. SSD's effectiveness surpasses that of traditional two-shot detectors, representing a notable departure from multi-stage object detection methods. Its versatility extends to various computer vision tasks, significantly bolstering efficiency and speed, crucial for real-time applications requiring swift decision-making based on object recognition within a scene.

In the proposed system, a centroid tracker is integrated to ensure consistent and precise evaluation of objects identified by the SSD. Computation of centroids within bounding boxes enhances tracking accuracy and facilitates geographic localization of objects. By incorporating a centroid tracker alongside the SSD, the system calculates centroids to encapsulate and monitor identifiable objects across frames. Each object is assigned a unique identification (ID) by the tracker, enabling independent tracking and consistent recording of object movement over time. This unique ID system streamlines efficient tracking of varying numbers of objects concurrently by maintaining the correlation between the ID and the object, ensuring uninterrupted record-keeping and simplifying the process of reconstructing object trajectories. Consequently, this enhances overall tracking accuracy and system efficiency.

The permanence of the identification number within this system significantly enhances the tracking process, facilitating easier reconstruction of object movements over time. This permanence ensures accuracy and reliability, further fortifying the system's ability to track and analyze object trajectories effectively. The seamless integration of SSD with a centroid tracker signifies a substantial advancement in computer vision technology, promising unparalleled precision and efficiency in real-time object detection and tracking. This innovative approach not only meets the demands of contemporary applications but also lays the groundwork for future advancements in computer vision, paving the way for enhanced capabilities and functionalities across various domains.

5.2 Future Scope

1. Addressing Energy Waste:

The research aims to combat the prevalent issue of unregulated electrical energy consumption, particularly in communal settings like restrooms and public spaces, where lighting fixtures often remain active unnecessarily, even in unoccupied areas.

2. Real-time Human Counter System:

Introducing a cutting-edge real-time human counter system designed to optimize energy efficiency by accurately tracking human presence and adjusting lighting accordingly, thus curbing excessive electricity usage and promoting sustainability.

3. Development of Open-Source Framework:

Utilizing the TensorFlow platform to develop an open-source framework, providing a user-friendly interface for creating sophisticated object-counting systems tailored to diverse environments.

4. Revolutionizing Lighting Management:

Leveraging the advanced capabilities of the proposed system in object recognition, tracking, and counting to revolutionize the management of lighting systems in public spaces, leading to significant energy conservation and operational efficiency.

5. Balancing Functionality and Sustainability:

Ensuring optimal lighting conditions while minimizing energy consumption by accurately identifying human presence and monitoring movements in public areas, thereby striking a balance between functionality and sustainability.

6. Beyond Energy Reduction:

Going beyond mere energy reduction, the research seeks to pioneer a unique approach that not only mitigates energy usage but also reduces associated costs, contributing to both environmental and economic sustainability.

7. Comprehensive Analysis:

Conducting a thorough analysis of the real-time human counting system, delving into the intricate mechanisms of object recognition and tracking to assess its potential implications for energy conservation in diverse public contexts.

8. Integration with Lighting Management Techniques:

Examining the integration of the proposed system with various lighting management techniques to evaluate its effectiveness in conserving energy and optimizing lighting conditions across different settings.

9. Insights for Informed Decision-Making:

Providing comprehensive insights into the implications of leveraging advanced technologies in lighting management systems to facilitate informed decision-making and proactive measures towards achieving sustainable energy practices.

10. Towards Sustainable Energy Practices:

Through concerted research efforts and proactive measures, striving towards realizing a future where energy conservation is not just an aspiration but a tangible reality, contributing to global efforts towards environmental sustainability.

REFERENCES

1. R. Roslina, A. Amelia, H. Pranoto, and B. V. Sundawa, “System of smart detection and control to electrical energy for saving of electrical energy consumption,” 2021, pp. 2454–2465.
2. H. Mokayed, T. Z. Quan, L. Alkhaled, and V. Sivakumar, “Real-time human detection and counting system using deep learning computer vision techniques,” in Artificial Intelligence and Applications, vol. 1, no. 4, 2023, pp. 221–229.
3. S. Chun and C.-S. Lee, “Applications of human motion tracking: Smart lighting control,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2013.
4. J. Liu, Y. Liu, Y. Cui, and Y. Q. Chen, “Real-time human detection and tracking in complex environments using single RGBD camera,” in 2013 IEEE International Conference on Image Processing, 2013, pp. 3088–3092.
5. M. Elsisi, M.-Q. Tran, K. Mahmoud, M. Lehtonen, and M. M. Darwish, “Deep learning-based industry 4.0 and internet of things towards effective energy management for smart buildings,” Sensors, vol. 21, no. 4, p. 1038, 2021.
6. J. Redmon and et al., “You only look once: Unified, real-time object detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
7. S. Ahmadi-Karvigh, A. Ghahramani, B. Becerik-Gerber, and L. Soibelman, “Real-time activity recognition for energy efficiency in buildings,” Applied Energy, vol. 211, pp. 146–160, 2018.
8. W. Liu and et al., “SSD: Single shot multi-box detector,” in Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I, vol. 14. Springer International Publishing, 2016.
9. K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” in ´ Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2961–2969.
10. R. Girshick, “Fast R-CNN,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1440–1448.
11. T. Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, ´ “Feature pyramid networks for object detection,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125.
12. Y. He, C. Zhu, J. Wang, M. Savvides, and X. Zhang, “Bounding box regression with uncertainty for accurate object detection,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2888–2897.
13. W. Abd-Almageed, M. Hussein, M. Abdelkader, and L. Davis, “Realtime human detection and tracking from mobile vehicles,” in 2007 IEEE Intelligent Transportation Systems Conference, 2007, pp. 149–154.
14. S. Lu, C. Hameen, and A. Aziz, “Dynamic HVAC operations with real-time vision-based occupant recognition system,” in 2018 ASHRAE Winter Conference, Chicago, 2018.
15. O. Mata, J. I. Mendez, P. Ponce, T. Peffer, A. Meier, and A. Molina, ´ “Energy savings in buildings based on image depth sensors for human activity recognition,” 2023, p. 1078.
16. A. P. Chandrasiri and D. Geekiyanage, “Real-time object detection system for building energy conservation: An IP camera based system,” THIRTY-FOURTH ANNUAL CONFERENCE, 2018.
17. R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza, and J. Almazan, ´ “Expert video-

surveillance system for real-time detection of suspicious behaviors in shopping malls," Expert Systems with Applications, vol. 42, no. 21, pp. 7991–8005, 2015.
18. S. V. Viraktamath, M. Katti, A. Khatawkar, and P. Kulkarni, "Face detection and tracking using OpenCV," The SIJ Transactions on Computer Networks and Communication Engineering (CNCE), vol. 1, no. 3, pp. 45–50, 2013.
19. A. Pennisi, D. D. Bloisi, and L. Iocchi, "Online real-time crowd behavior detection in video sequences," Computer Vision and Image Understanding, vol. 144, pp. 166–176, 2016.
20. C. Ning, H. Zhou, Y. Song, and J. Tang, "Inception single shot multibox detector for object detection," in 2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), 2017, pp. 54

APPENDIX A

Human Detection to Mitigate Excessive Electrical Energy Usage in Communal Spaces

Aabhas Bisaria
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
aabhasbisaria@gmail.com

Devesh Chauhan
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
deveshc2002@gmail.com

Harsh Varshney
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
harshjul11@gmail.com

Sushil Kumar
DSDL, Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
drsushil.cs@gmail.com

Umang Rastogi
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
rastogi225103@gmail.com

Vineet Sharma
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
vineet.sharma@kiet.edu

***Abstract*—The purpose of this study is to investigate the issue of increased electrical energy consumption in communal settings, with a specific focus on areas such as restrooms and multiplexes as the primary areas of investigation. Not only does the observed behavior increase the financial requirements of end-users and facility operators, but it also places a major strain on energy grids, which in turn leads to a rise in carbon emissions and an intensification of environmental concerns. A unique approach that makes use of the TensorFlow Object Counting API is proposed by the current inquiry as a means of addressing the concern that was highlighted earlier. The purpose of this strategy is to develop a system that can precisely count human subjects in real-time when they are present. To effectively decrease the amount of electricity that is not required, the primary objective of this cutting-edge system is to precisely identify and count the number of people that are present in shared spaces. When it comes to the development of advanced object counting systems, the open-source framework that has been developed on the TensorFlow is a platform that offers a method that is both user-friendly and efficient. Within the realms of item identification, tracking, and counting, the system exhibits capabilities that are both spectacular and resilient. By addressing the dual problems of rising electricity bills and negative effects on the environment, this technology offers a potential solution to the problem. Through this study, an attempt is made to investigate the factors that lead to excessive energy use and to propose a practical solution to this significant problem.**

***Index Terms*—Electrical Energy Saving, Human Detection, Single Shot Detector, Object Tracking Algorithm**

I. Introduction

In the current setting, one of the most important criteria that must be satisfied is to make the most of the available resources and to spend them as efficiently as possible. Currently, the problem at hand is the unregulated and ongoing consumption of electrical energy [1], which frequently surpasses the requirements that are being met. Because it raises a considerable concern, this constitutes a significant problem that calls for an immediate solution. There is a wide variety of various communal settings in which the observed predominance of wastefulness is apparent. Restrooms, public restrooms, and multiplexes are some examples of these types of locations; however, this list is not exhaustive. The analysis of the data reveals a discernible pattern in which the lighting fixtures located within the designated areas consistently tend to remain activated, even when there are no occupants present. This is particularly obvious in situations where there are no people around. Even when there are no other individuals in the vicinity, this is still the case.

To solve the issue of increased energy consumption, the primary objective of the current research is to offer a real-time human counter system as a means of tackling the problem [2]. The current system has been painstakingly constructed with the idea of accomplishing the primary objective of optimizing energy efficiency through the removal of superfluous electricity usage [3]. This target has been the fundamental inspiration for the construction of the system. When it comes to creativity and originality, the open-source framework that is currently being evaluated demonstrates an impressive amount of exceptional qualities. To construct this framework, the TensorFlow platform was utilized. The system provides users with a methodology that is not only user-friendly but also very efficient. This is done to simplify the creation of sophisticated object-counting systems. The system that is the focus of this investigation has an outstanding level of adaptability and efficiency in its operation. The resilient capabilities of the system in the areas of object recognition, tracking, and counting may be principally responsible for this particular consequence [4]. Taking into consideration the possibilities of this technology, there is a significant likelihood that it will bring about a revolutionary change in the management and operation of lighting systems in public spaces.

The technology that is being investigated demonstrates promising capabilities in properly recognizing the presence of individuals in public places and accurately monitoring the movements of those individuals. It is being taken into mind that certain skills are available. In addition to this, it will ensure that the lighting conditions within these specific regions are not only adequate but also excellent in terms of the amount of energy that they consume. When it comes to the study that we are conducting, our major purpose is not simply to cut down on the amount of electricity that is used [5]. The methodology that has been shown may be a unique approach that has the potential to result in considerable reductions in both the amount of energy that is used and the costs that are associated with both of these things.

A thorough analysis of our real-time human counting system is going to be carried out in the subsequent sections, which will thereafter be presented [6]. It will be important to perform a comprehensive examination of the technological components to properly execute the endeavor that has been recommended. This investigation will encompass the intricate mechanisms that are involved in the process of object recognition and tracking. As an additional point of interest, the primary objective of this research endeavor is to conduct a comprehensive examination into the combination of the system that was mentioned earlier with various techniques for the management of lighting. A thorough examination will also be carried out to evaluate the possible implications of this technology concerning the conservation of electricity in a range of public contexts [7]. This investigation will be carried out to examine the potential implications of this technology. The fundamental objective of the current study, which intends to provide in-depth information on these implications, is to provide a thorough understanding of the larger implications of our research in effectively addressing the existing challenges related to energy conservation. This is the primary objective of the study.

II. Related Works

The discipline of computer vision has recognized the significance of object detection, which is a key component that possesses tremendous utility across a wide variety of applications. There has been a noticeable shift toward the growth of real-time detection systems over the past ten years.The Single Shot Detector (SSD), first introduced in [8], represents a significant departure from traditional two-shot detectors such as R-CNN. The current study presents a new strategy that provides a methodology that can be used more effectively. By implementing a unified strategy that simultaneously combines regional proposal generation and object detection, the method known as SSD is able to achieve its goal. As a result of the strategy, an important paradigm shift took place in the respective field, especially in situations where convenience and efficiency are extremely important. One of the most innovative approaches to object recognition in computer vision was presented in the research paper ”You Only look Once: Unified, Real-Time Object Detection”. [6]. By adopting a unified model that analyzes the entire image at once, the work addressed the limitations of existing methods. The article was published in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. The success created a completely new paradigm for real-time object detection, marking a major departure from previous multi-pass methods. One of the main highlights is the You Only Look Once (YOLO) algorithm proposed in the paper. The image was first segmented into a grid, and then bounding boxes and class probabilities were predicted for each grid cell. This unified method eliminated the requirement for separate object location and classification, resulting in both faster and more efficient object detection. The simplicity and efficiency of the YOLO architecture was the basis for further research in the field.

One of the most innovative approaches to object identification in computer vision was provided in the research article titled “You Only Look Once: Unified, Real-time Object Detection” [6]. Through the introduction of a unified model that analyzed the full image in a single pass, the work addressed the constraints of existing methods. The article was published in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. As a result of the success, a completely new paradigm for real-time object detection was created, which marked a significant departure from the previously dominant multi-pass methodology. One of the most important highlights is the You Only Look Once (YOLO) algorithm that was proposed in the paper. The image was first segmented into a grid and then bounding boxes and class probabilities were predicted for each grid cell. This unified method eliminated the requirement for separate object location and classification, resulting in both faster and more efficient object detection. The simplicity and efficiency of the YOLO architecture was the basis for further research in the field.

A significant advancement in computer vision, particularly in instance segmentation, is demonstrated by the study ”Mask R-CNN” [9]. The Mask R-CNN model, based on the Faster R-CNN framework, presents a novel architecture that can predict object bounding boxes, class labels, and pixel-level masks simultaneously. This research paper examines the most important achievements, approaches and implications of the mask R-CNN model in the context of computer vision. The main contribution of the study is the ability of the Mask R-CNN paper to perform instance segmentation by extending the faster R-CNN architecture. In addition to the branches created for class classification and bounding box detection, this model also have a parallel branch for segmentation mask prediction. This technique made it possible to precisely define object boundaries at the pixel level, which was important for applications that require precise object placement and separation. In terms of integrating pixel-level mask prediction into an object detection framework, the innovation of Mask R-CNN marked a departure from traditional approaches that were used in the past. The challenges of accurately understanding spatial relationships through the simultaneous prediction of masks, jump boxes and class labels showed how versatile the model was. Subsequent research and applications in several fields have made Mask R-CNN a pixel-by-pixel segmentation industry standards.

One of the most important contributions to the field of object identification is the research article titled “Fast R-CNN” [10]. A unified framework that merged region proposal generation and object identification into a single model was established by Fast R-CNN to address the computational inefficiencies that were present in its predecessors. Girshick’s Fast R-CNN has had a notable impact on the field of object detection, inspiring following research and becoming a reference point for new algorithms. This research review examines the key principles, contributions, and impact of Girshick’s Fast R-CNN. As a result of the model’s potency in simultaneously handling region proposal and object detection tasks, more evolution have been inspired, such as Faster R-CNN and later versions. There are applications for Fast R-CNN in fields such as surveillance, automatic systems, and image-based searching application, which means that its influence goes beyond the realm of academics.

Feature Pyramid Networks for Object Detection address the challenge of differences in the field of object recognition [11]. After the authors realized how critical it is to detect features by capturing them at a variety of scales, they came up with the concept for the Feature Pyramid Network (FPN), which is an initialism for the framework. The most significant contributions, methodology, and impacts of FPN are investigated in this literature review, which focuses on object detection as its primary topic of investigation. The most significant addition that the study has made is the presentation of the design for the Feature Pyramid Network, which should be examined first and foremost. The formation of a feature pyramid, which is then employed to address the difficulty of scale variance, is accomplished by the utilization of a single convolutional network. Each of these feature maps gathers information at a specific scale, and their resolutions range from low to high. Features are the building blocks of this pyramid. The design, which enables high-level semantic properties to be paired with exact geographical information, makes it feasible to perform object recognition that is both more accurate and scalable. This functionality is made possible by the architecture.

III. Proposed System

In the system that is currently under investigation, two counting functions are considered to be of critical importance. The power to quantify the influx of individuals entering a certain location and the capability to monitor the count of individuals leaving that same area are both examples of these functionalities. To properly achieve the intended outcome, it is vitally necessary to focus one’s attention solely on the persons who have passed the threshold of intrusion. This is a requirement that cannot be avoided. By establishing the direction of the movement, specifically, whether it is oriented in a downward or upward trajectory, it is possible to make the completion of this aim easier to execute. By utilizing bounding boxes concept, which enable a more extensive visual representation, it is possible to generate a visual depiction of the detection of human. These bounding boxes [12] try to attempt enclose the human items that have been recognized as belonging to these categories. Additionally, a one-of-a-kind identification number is attached to each bounding box. This number serves the purpose of indicating the order in which the human artifacts were found. The architecture of the system has been painstakingly designed by the specifications through the use of a cascade method. The methodology that has been proposed begins with the identification of a human entity, which is then followed by the monitoring and tracking of that individual throughout a series of frames that are located close to one another. A high level of proficiency in the execution of counting precisely and tracking a large number of items at the same time is demonstrated by the methodology that has been proposed. Additionally, it can efficiently cope with situations in which a human-made thing that has never been seen before comes into view, and it immediately assigns it a distinct identity. This is a capability that it possesses.

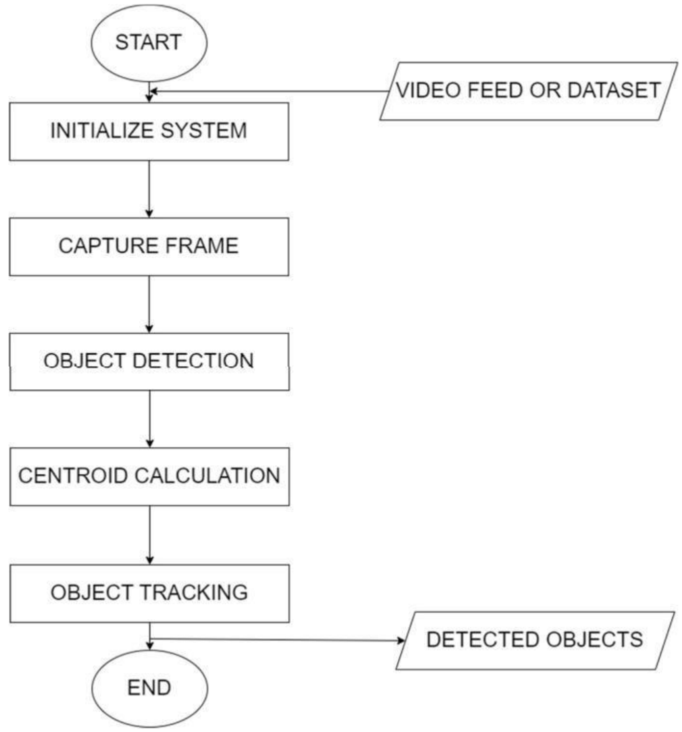


Fig. 1. Flow chart of proposed methodology

Therefore, to ensure the proper execution of this system, it is necessary to incorporate two essential components, which are as follows: both a human detection [13] module that is only devoted to the identification of human beings and a tracking mechanism that is exclusively designed to monitor the human objects that have been recognized are included in this system.

IV. Methodology

By utilizing a computer vision [14] investigation that is centered around the utilization of the Single Shot Detector (SSD) in conjunction with the Mobile Net architecture for real-time object detection [15], the researchers have used a computer vision investigation in the defining of the project. One of the most important feature of improving this functionality of smartphone devices, IP cameras, and scanners is

Algorithm 1 Object Tracking Algorithm

```
Initialize:
- List of detected objects
- Dictionary to store object IDs and their respective centroids
function DETECT_OBJECTS(frame)
    # Utilize Single Shot Detector (SSD) to detect objects
    detected_objects ← SSD.detect(frame)
    return detected_objects
end function
function TRACK_OBJECTS(detected_objects)
    for each object in detected_objects do
        if object is new then
            Assign a unique ID
            Add object ID and centroid to dictionary
        else
            Update centroid for existing object
        end if
    end for
end function
function DISPLAY_TRACKING_RESULTS(frame)
    for each object ID and centroid in dictionary do
        Draw bounding box around object using centroid
        Display object ID and bounding box on frame
    end for
end function
Main Loop:
while video is running do
    Capture frame from video
    detected_objects ← DETECT_OBJECTS(frame)
    TRACK_OBJECTS(detected_objects)
    DISPLAY_TRACKING_RESULTS(frame)
end while
```

Fig. 2. Pseudo code of object tracking algorithm

the establishment of this cutting-edge amalgamation, which efficiently optimizes both speed and efficiency demonstrated in Figure 3. In contrast to two-shot detectors like R-CNN, SSD can identify objects in a single iteration, which is a significant advantage over the latter. A sort of object detection [16] model known as the SSD is distinguished from two-shot detectors by the strategy that it takes. In contrast to two-shot detectors, which normally carry out the duties of region proposal and object identification in consecutive processes, SSD is capable of carrying out both functions concurrently. This indicates that SSD is capable of generating region proposals and detecting objects simultaneously over the network, which ultimately results in faster processing that is both accurate and precise. Real-time object detection [17] is made possible in a various of applications thanks to this approach, which removes the need for additional computing stages. As opposed to two shot detectors like R-CNN, the efficiency of SSD is remarkably higher. An important departure from the traditional multi-stage object detection methods is represented by the employment of SSD. By integrating region proposal and object detection, which is made possible by the utilization of SSD, the system increases both its efficiency and its speed. The ability to make decisions quickly, based on the recognition of objects inside a given scene, is required for real-time applications for them to function well. Centroid tracker is a well-established method that is used to track objects across multiple frames.

Fig. 3. Stock video frame without centroid allocation

It is essential to make use of this tracking technique [18] to maintain regularity and difference over consecutive frames to guarantee a thorough evaluation of the items selected by SSD. The determination of the centroid of a bounding box is considered to be the core idea that is at play here. It is possible to notably increase the accuracy of object detection and tracking in visual images by establishing this computational strategy. A centroid tracker is used to determine the centroid of the bounding boxes that are produced by the Single Shot Multi-Box Detector (SSMD). To sum up and depict items that have been successfully detected inside specific frames of a given dataset or video sequence [19], bounding boxes inclusive of a set of two-dimensional coordinates, which are denoted as (x, y). Bounding boxes serve the aim of encapsulating and outlining things. To find the centroid of the bounding boxes, the tracker technique makes use of a computational approach. This makes it possible to discover the point within the frame that is the most central to the item. By making use of the concept of an object's centroid, it is possible to achieve the competence of monitoring and determining the trajectory and spatial coordinates of an object across number of video frames.

To ensure that the Centroid Tracker can perform its functions correctly and effectively, every object that is being observed must be assigned a unique and customized identity demonstrated in Figure 4. This enables the system to monitor and record the object's spatial displacement over consecutive frames. The tracking mechanism of the system is dependent on the unique identification (ID) [20] that is issued to each object. As a result of the system's consistent establishment of a correlation between the unique identification and the object in question, it can monitor and record the trajectory of the object throughout the procedure. In circumstances that involve several different items, the fact that it can independently track distinct things and provide a unique identity for each of them makes it extremely significant. The numerical IDs that have been provided are permanent, which ensures that continuity will be maintained and makes the process of reconstructing the line of travel of the object easier to accomplish.

Fig. 4. Stock video frame with centroid allocation through proposed methodology

V. Results Analysis

To accomplish the goals of this investigation, the system that is being studied is put through a series of tests on the graphics processing unit (GPU) of a computer that is located in the investigation area. The dataset that is being utilized in these experiments has been selected with considerable care. The central processing unit (CPU) was placed in the role of the target machine during the initial testing of the system. Regardless of this, it was quickly rejected because GPU displayed substantially higher levels of performance than previously thought. Only the number of line incursions going up and down are significant for counting. Tables I and II display the accurate manual count and the predicted count using the proposed technique for various camera angles. To get the final accuracy percentage, calculating the frequency of error by subtracting the anticipated count from the observed count. For the more accurate normal model, the ultimate accuracy is 91.66%, while for the faster model, it is 78.33%. Both results are elaborately outlined and shown separately below.

Rate of error = sum of frequency of error/(sum of anticipated increase and anticipated decrease) × 100%
= (0+1+1+1+2)/(1+4+6+3+11+11+2+6+2+4)× 100%
= (5/50) × 10 %
= 0.1000 × 100%
= 10%
Overall accuracy = total anticipated count - rate of error
= 100% - 10%
= 90%

Rate of error = sum of frequency of error/(sum of anticipated increase and anticipated decrease)× 100%
= (0+5+3+0+1)/(1+4+6+3+11+11+2+6+2+4)× 100%
= (9/50)×100%
= 0.18×100%
= 18%

Overall accuracy = total anticipated count - rate of error
= 100% - 18%
= 82%

TABLE I
Accuracy outcome of standard model

Serial No.	Camera positioning	Anticipated increase	Anticipated decrease	Actual increase	Actual decrease
I	Anterior vision	1	4	1	4
II	Anterior vision	6	3	7	3
III	Top vision	11	11	11	12
IV	Top vision	2	6	2	5
V	Top vision	2	4	1	3

TABLE II
Accuracy outcome with increased speed

Serial No.	Camera positioning	Anticipated increase	Anticipated decrease	Actual increase	Actual decrease
I	Anterior vision	1	4	2	3
II	Anterior vision	6	3	3	1
III	Top vision	11	11	10	9
IV	Top vision	2	6	2	6
V	Top vision	2	4	2	5

VI. Conclusion

The research embarked on a computer vision study aimed at leveraging the Single Shot Detector (SSD) approach alongside the Mobile Net architecture to enable real-time object detection. This integration significantly enhances speed and efficiency crucial for mobile devices, IP cameras, and scanners. Unlike two-shot detectors like R-CNN, SSD can identify objects in a single iteration, combining region proposal and object identification into a unified phase, eliminating the need for a separate region proposal network (RPN). This consolidation streamlines processing, optimizes resources and simplifies the detection pipeline. SSD's efficacy surpasses that of traditional two-shot detectors like R-CNN, offering a notable departure from multi-stage object detection techniques. Its benefits extend to various computer vision tasks, enhancing both efficiency and speed, crucial for real-time applications requiring swift decision-making based on object recognition within a scene.

This proposed system integrate a centroid tracker for object tracking across successive frames, ensuring consistency and clarity in evaluating objects identified by the Single Shot Detector (SSD). Centroid computation within bounding boxes enhances tracking accuracy and geographic localization of objects. By utilizing a centroid tracker alongside the SSD, the system computes centroids to encapsulate and track identifiable objects across frames. Each object is assigned a unique identification (ID) by the tracker, facilitating independent tracking and consistent recording of object movement over time. This unique identity system enables efficient tracking of various number of objects simultaneously by maintaining relationship between the unique identification number and the object, make it certify that it maintain uninterrupted record and simplifies the process for redeveloping object trajectories. This permanent nature of the identification number is streamlined for the tracking process, making it accurate easier to reconstruct object movements over time.

REFERENCES

[1] R. Roslina, A. Amelia, H. Pranoto, and B. V. Sundawa, "System of smart detection and control to electrical energy for saving of electrical energy consumption," 2021, pp. 2454–2465.

[2] H. Mokayed, T. Z. Quan, L. Alkhaled, and V. Sivakumar, "Real-time human detection and counting system using deep learning computer vision techniques," in *Artificial Intelligence and Applications*, vol. 1, no. 4, 2023, pp. 221–229.

[3] S. Chun and C.-S. Lee, "Applications of human motion tracking: Smart lighting control," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2013.

[4] J. Liu, Y. Liu, Y. Cui, and Y. Q. Chen, "Real-time human detection and tracking in complex environments using single RGBD camera," in *2013 IEEE International Conference on Image Processing*, 2013, pp. 3088–3092.

[5] M. Elsisi, M.-Q. Tran, K. Mahmoud, M. Lehtonen, and M. M. Darwish, "Deep learning-based industry 4.0 and internet of things towards effective energy management for smart buildings," *Sensors*, vol. 21, no. 4, p. 1038, 2021.

[6] J. Redmon and et al., "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[7] S. Ahmadi-Karvigh, A. Ghahramani, B. Becerik-Gerber, and L. Soibelman, "Real-time activity recognition for energy efficiency in buildings," *Applied Energy*, vol. 211, pp. 146–160, 2018.

[8] W. Liu and et al., "SSD: Single shot multi-box detector," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I*, vol. 14. Springer International Publishing, 2016.

[9] K. He, G. Gkioxari, P. Dolla´r, and R. Girshick, "Mask R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.

[10] R. Girshick, "Fast R-CNN," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1440–1448.

[11] T. Y. Lin, P. Dolla´r, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2117–2125.

[12] Y. He, C. Zhu, J. Wang, M. Savvides, and X. Zhang, "Bounding box regression with uncertainty for accurate object detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2888–2897.

[13] W. Abd-Almageed, M. Hussein, M. Abdelkader, and L. Davis, "Real-time human detection and tracking from mobile vehicles," in *2007 IEEE Intelligent Transportation Systems Conference*, 2007, pp. 149–154.

[14] S. Lu, C. Hameen, and A. Aziz, "Dynamic HVAC operations with real-time vision-based occupant recognition system," in *2018 ASHRAE Winter Conference*, Chicago, 2018.

[15] O. Mata, J. I. Me´ndez, P. Ponce, T. Peffer, A. Meier, and A. Molina, "Energy savings in buildings based on image depth sensors for human activity recognition," 2023, p. 1078.

[16] A. P. Chandrasiri and D. Geekiyanage, "Real-time object detection system for building energy conservation: An IP camera based system," *THIRTY-FOURTH ANNUAL CONFERENCE*, 2018.

[17] R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza, and J. Almaza´n, "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7991–8005, 2015.

[18] S. V. Viraktamath, M. Katti, A. Khatawkar, and P. Kulkarni, "Face detection and tracking using OpenCV," *The SIJ Transactions on Computer Networks and Communication Engineering (CNCE)*, vol. 1, no. 3, pp. 45–50, 2013.

[19] A. Pennisi, D. D. Bloisi, and L. Iocchi, "Online real-time crowd behavior detection in video sequences," *Computer Vision and Image Understanding*, vol. 144, pp. 166–176, 2016.

[20] C. Ning, H. Zhou, Y. Song, and J. Tang, "Inception single shot multibox detector for object detection," in *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2017, pp. 549–554.

APPENDIX B

Title of Paper: Human Detection to Mitigate Excessive Electrical Energy Usage in Communal Spaces

Name of IEEE Conference: International Conference on Intelligent Systems for Cybersecurity (ISCS 2024) organised by The North Cap University, Gurugram

Date of Submission: 29th February 2024

Proof:

International Conference on Intelligent Systems for Cybersecurity (ISCS 2024) organised by The NorthCap University, Gurugram : Submission (323) has been created. Inbox

Microsoft CMT 29 Feb
to me

Hello,

The following submission has been created.

Track Name: International Conference on Intelligent Systems for Cybersecurity (ISCS 2024)

Paper ID: 323

Paper Title: Human Detection to Mitigate Excessive Electrical Energy Usage in Communal Spaces

Abstract:
The purpose of this study is to investigate the issue of increased electrical energy consumption in communal settings, with a specific focus on areas such as restrooms and multipl[redacted]rimary

Date of Acceptance: 16th April 2024

Proof:

Fwd: Acceptance Notification for International Conference on Intelligent Systems for Cybersecurity (ISCS 2024) organised by The NorthCap University, Gurugram - Paper ID: 323 Inbox

d Dr. Sushil Ku... 11:30 AM
to me

---------- Forwarded message ---------
From: **Microsoft CMT** <email@msr-cmt.org>
Date: Tue, 16 Apr 2024 at 10:45
Subject: Acceptance Notification for International Conference on Intelligent Systems for Cybersecurity (ISCS 2024) organised by The NorthCap University, Gurugram - Paper ID: 323
To: Sushil Kumar <drsushil.cs@gmail.com>

Dear Sushil Kumar,

We are pleased to inform you that your paper with id "323", titled "Human Detection to Mitigate Excessive Electrical Energy Usage in Communal Spaces" has been accepted for presentation at the "International Conference on Intelligent Systems for Cybersecurity (ISCS 2024) organised by The NorthCap University, Gurugram", scheduled to take place between 5/2/2024 - 5/3/2024 in collaboration with IEEE. The proceedings of ISCS 2024 will be forwarded to be published in the IEEE Xplore, the digital library of IEEE which is currently indexed in SCOPUS, Web of Science, etc.

Your paper underwent a rigorous review process, and the reviewers were impressed by the quality of your work and its relevance to the conference themes. Congratulations on this achievement!

Kindly ensure the following points before uploading the final paper:

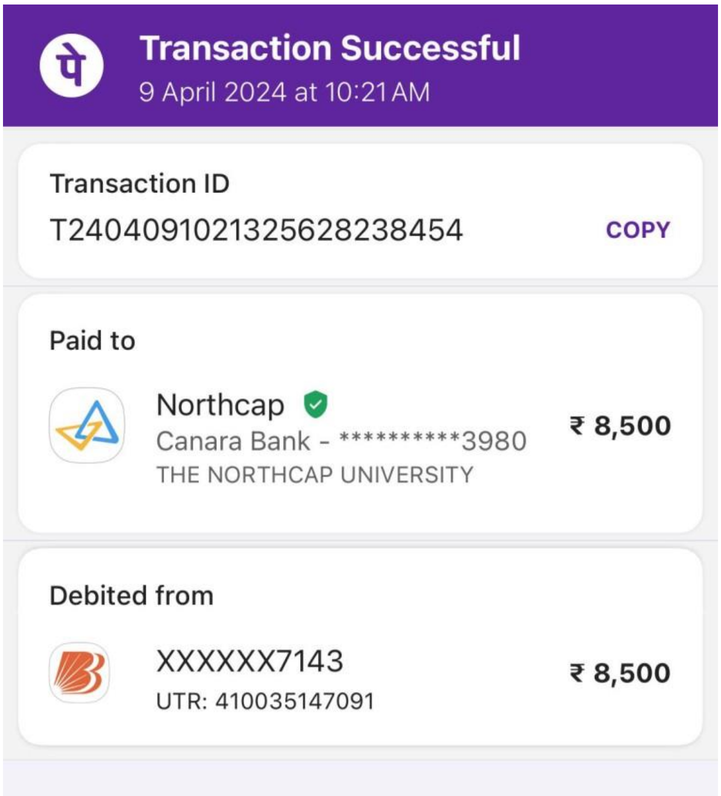
1) The format must be as per IEEE Template https://www.ieee.org/conferences/publishing/templates.html with maximum 6 pages.

2) Minimum 20 references must be in the paper and all references must be cited in the text. Like [1], [2] ...

3) Carefully look at the typographical/grammatical errors in the paper.

Date of Registration: 9th April 2024

Proof:



Date of Conference: 4th May 2024

Proof:

THE NORTHCAP UNIVERSITY

Technical Sessions – Day 2

Session9: **Day and Date: Saturday, 4th May 2024**
Venue: Room No N208 **Timing: 9:30 am – 1:00 pm**
Track: AI and Machine Learning for Cybersecurity
Session Chair: Dr. Neetu Sardana, Dr. Deepika Varshney
Session Coordinator: Dr. Tamalika
Contact Details: 7838976484
MS Teams Link: https://teams.live.com/meet/9516621850218?p=U69v0TWAlfjUd0Gq

S. No.	Time	Paper Id	Authors	Paper Title
10	11:00 AM - 11:10 AM	246	Gill, Kanwarpartap Singh*; Agarwal, Muskan; Upadhyay, Deepak; Devliyal, Swati	From Pixels to Insights: Harnessing Deep Learning for Accurate Plant Pathology Diagnosis
11	11:10 AM - 11:20 AM	254	Gill, Kanwarpartap Singh*; Mittal, Khushi; Chattopadhyay, Saumitra; Singh, Mukesh	Enhancing Coral Health Evaluation with VGG16: CNN - Powered Approach for Streamlined Reef Surveillance and Preservation
12	11:20 AM-11:30 AM	256	Gill, Kanwarpartap Singh*; Agarwal, Muskan; Thapliyal, Nitin; Rawat, Ramesh Singh	Savouring Tradition with a Technological Twist: Pandang Cuisine Classification using Transfer Learning
13	11:30 AM -11:40 AM	259	Gill, Kanwarpartap Singh*; Mittal, Khushi; Upadhyay, Deepak; Devliyal, Swati	Trailblazing Strategies for Solar Panel Maintenance: Employing VGG19 for Early Detection of Damage
14	11:40 AM -11:50 AM	260	Gill, Kanwarpartap Singh*; Singla, Muskan; Malhotra, Sonal; Devliyal, Swati	Innovative Methodology Utilizing Vibration Sensor Data for Early Detection and Prediction in Wind Turbine Systems
15	11:50 AM - 12:00 PM	323	Bisaria, Aabhas; Chauhan, Devesh; Varshney, Harsh; Kumar, Sushil*; RASTOGI, UMANG; Sharma, Vineet	Human Detection to Mitigate Excessive Electrical Energy Usage in Communal Spaces

Human Count SK

by HIMANSHU CHAUDHARY

Submission date: 29-Feb-2024 09:01AM (UTC+0530)
Submission ID: 2288463622
File name: manuscript.pdf (638.9K)
Word count: 4277
Character count: 23807

Human Detection to Mitigate Excessive Electrical Energy Usage in Communal Spaces

Aabhas Bisaria
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
aabhasbisaria@gmail.com

Devesh Chauhan
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
deveshc2002@gmail.com

Harsh Varshney
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
harshjul11@gmail.com

Sushil Kumar
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
drsushil.cs@gmail.com

Name
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
write email id

Name
Computer Science and Engineering
KIET Group of Institutions, Delhi-NCR
Ghaziabad, India
Write email id

***Abstract*—The purpose of this study is to investigate the issue of increased electrical energy consumption in communal settings, with a specific focus on areas such as restrooms and multiplexes as the primary areas of investigation. Not only does the observed behavior increase the financial requirements of end-users and facility operators, but it also places a major strain on energy grids, which in turn leads to a rise in carbon emissions and an intensification of environmental concerns. A unique approach that makes use of the TensorFlow Object Counting API is proposed by the current inquiry as a means of addressing the concern that was highlighted earlier. The purpose of this strategy is to develop a system that can precisely count human subjects in real-time when they are present. To effectively reduce the amount of electricity that is not required, the primary objective of this cutting-edge system is to precisely identify and count the number of people that are present in shared spaces. When it comes to the development of advanced object counting systems, the open-source framework that has been developed on the TensorFlow platform offers a method that is both user-friendly and efficient. Within the realms of item identification, tracking, and counting, the system exhibits capabilities that are both spectacular and resilient. By addressing the dual problems of rising electricity bills and negative effects on the environment, this technology offers a potential solution to the problem. Through this study, an attempt is made to investigate the factors that lead to excessive energy use and to propose a practical solution to this significant problem.**

***Index Terms*—Electrical Energy Saving, Human Detection, Single Shot Detector, Object Tracking Algorithm**

I. Introduction

In the current setting, one of the most important criteria that must be satisfied is to make the most of the available resources and to spend them as efficiently as possible. Currently, the problem at hand is the unregulated and ongoing consumption of electrical energy, which frequently surpasses the requirements that are being met. Because it raises a considerable concern, this constitutes a significant problem that calls for an immediate solution. There is a wide variety of various communal settings in which the observed predominance of wastefulness is apparent. Restrooms, public restrooms, and multiplexes are some examples of these types of locations; however, this list is not exhaustive. The analysis of the data reveals a discernible pattern in which the lighting fixtures located within the designated areas consistently tend to remain activated, even when there are no occupants present. This is particularly obvious in situations where there are no people around. Even when there are no other individuals in the vicinity, this is still the case.

To solve the issue of increased energy consumption, the primary objective of the current research is to offer a real-time human counter system as a means of tackling the problem. The current system has been painstakingly constructed with the idea of accomplishing the primary objective of optimizing energy efficiency through the removal of superfluous electricity usage [1]. This target has been the fundamental inspiration for the construction of the system. When it comes to creativity and originality, the open-source framework that is currently being evaluated demonstrates an impressive amount of exceptional qualities. To construct this framework, the TensorFlow platform was utilized. The system provides users with a methodology that is not only user-friendly but also very efficient. This is done to simplify the creation of sophisticated object-counting systems. The system that is the focus of this investigation has an outstanding level of adaptability and efficiency in its operation. The resilient capabilities of the system in the areas of object recognition, tracking, and counting may be principally responsible for this particular consequence [2]. Taking into consideration the possibilities of this technology, there is a significant likelihood that it will bring about a revolutionary change in the management and operation of lighting systems in public spaces.

The technology that is being investigated demonstrates promising capabilities in properly recognizing the presence of individuals in public places and accurately monitoring the movements of those individuals. It is being taken into mind that certain skills are available. In addition to this, it will ensure that the lighting conditions within these specific regions are not only adequate but also excellent in terms of the amount of energy that they consume. When it comes to the study that we are conducting, our major purpose is not simply to cut down on the amount of electricity that is used [3]. The methodology that has been shown may be a unique approach that has the potential to result in considerable reductions in both the amount of energy that is used and the costs that are associated with both of these things.

A thorough analysis of our real-time human counting system is going to be carried out in the subsequent sections, which will thereafter be presented [4]. It will be important to perform a comprehensive examination of the technological components to properly execute the endeavor that has been recommended. This investigation will encompass the intricate mechanisms that are involved in the process of object recognition and tracking. As an additional point of interest, the primary objective of this research endeavor is to conduct a comprehensive examination into the combination of the system that was mentioned earlier with various techniques for the management of lighting. A thorough examination will also be carried out to evaluate the possible implications of this technology concerning the conservation of electricity in a range of public contexts [5]. This investigation will be carried out to examine the potential implications of this technology. The fundamental objective of the current study, which intends to provide in-depth information on these implications, is to provide a thorough understanding of the larger implications of our research in effectively addressing the existing challenges related to energy conservation. This is the primary objective of the study.

II. Related Works

The discipline of computer vision has recognized the significance of object detection, which is a key component that possesses tremendous utility across a wide variety of applications [4]. There has been a perceptible shift toward the growth of real-time detection systems over the past ten years. The Single Shot Detector (SSD), which was first presented in [4], represents a considerable departure from traditional two-shot detectors such as R-CNN. The current investigation presents a fresh strategy that provides a methodology that has the potential to be more efficiently utilized. Through the implementation of a unified strategy that combines the production of region suggestions and the detection of objects simultaneously in a single pass, the methodology known as SSD can accomplish its goals. A significant paradigm shift has occurred within the relevant domain as a result of the strategy, particularly in situations when expediency and effectiveness are of the utmost significance.

One of the most innovative approaches to object identification in computer vision was provided in the research article titled "You Only Look Once: Unified, Real-time Object Detection" [2]. Through the introduction of a unified model that analyzed the full image in a single pass, the work addressed the constraints of existing methods. The paper was published in the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. A whole new paradigm in real-time object detection was established as a result of the breakthrough, which marked a substantial departure from the multi-pass methodologies that were previously dominant.

One of the most important highlights is the You Only Look Once (YOLO) algorithm that was proposed in the paper. The image was first segmented into a grid by Redmon et al., and then bounding boxes and class probabilities were predicted for each grid cell. The elimination of the requirement for separate passes for object localization and classification was made possible by this unified method, which resulted in object detection that was both quicker and more effective. The straightforwardness and efficiency of the YOLO architecture served as the basis for later study in the sector.

A big step forward in the field of computer vision, notably in the area of instance segmentation, is represented by the research article titled "Mask R-CNN" [3]. A revolutionary architecture that is capable of simultaneously predicting object bounding boxes, class labels, and pixel-level masks is introduced by the Mask R-CNN model, which is built upon the Faster R-CNN framework. Within the context of the field of computer vision, this literature review investigates the most significant contributions, methodologies, and impacts that the Mask R-CNN model has achieved. The capacity of the Mask R-CNN article to execute instance segmentation by extending the Faster R-CNN architecture is the key contribution that the study makes. In addition to the branches that were already in place for bounding box detection and class classification, the model also included a parallel branch to predict segmentation masks. Using this method, it was possible to accurately delineate the boundaries of objects at the pixel level, which was a vital need in applications that required exact object localization and differentiation. When it came to incorporating pixel-level mask prediction into the object identification framework, the innovation that was presented by Mask R-CNN represented a break from the conventional methods that were previously used. The adaptability of the model was demonstrated in challenges that required a fine-grained comprehension of spatial relationships through the simultaneous prediction of masks, bounding boxes, and class labels. Since then, the pixel-wise segmentation capabilities of Mask R-CNN have established themselves as a standard in the industry, which has prompted future research and applications in a variety of fields, respectively.

One of the most important contributions to the field of object identification is the research article titled "Fast R-CNN" [4]. A unified framework that merged region proposal generation and object identification into a single model was established by Fast R-CNN to address the computational inefficiencies that were present in its predecessors. Girshick's Fast R-CNN has had a significant impact on the field of object identification, inspiring following research and becoming a reference point for benchmarking new algorithms. This literature review examines the key principles, contributions, and impact of Girshick's Fast

R-CNN. As a result of the model's effectiveness in simultaneously handling region proposal and object identification tasks, more developments have been inspired, such as Faster R-CNN and later versions. There are applications for Fast R-CNN in fields such as surveillance, autonomous systems, and image-based search, which means that its influence goes beyond the re4n of academics.

Feature Pyramid Networks for Object Detection address the challenge of scale variance in the field of object recognition [5]. After the authors realized how crucial it is to accurately detect features by capturing them at a variety of scales, they came up with the concept for the Feature Pyramid Network (FPN), which is an acronym for the framework. The most significant contributions, methodology, and impacts of FPN are investigated in this literature review, which focuses on object detection as its primary topic of investigation. The most significant addition that the study has made is the presentation of the design for the Feature Pyramid Network, which should be examined first and foremost. The formation of a feature pyramid, which is then employed to address the difficulty of scale variance, is accomplished by the utilization of a single convolutional network. Each of these feature maps gathers information at a specific scale, and their resolutions range from low to high. Features are the building blocks of this pyramid. The design, which enables high-level semantic properties to be paired with exact geographical information, makes it feasible to perform object recognition that is both more accurate and scalable. This functionality is made possible by the architecture.

III. Proposed System

In the system that is currently under investigation, two counting functions are considered to be of critical importance. The power to quantify the influx of individuals entering a certain location and the capability to monitor the count of individuals leaving that same area are both examples of these functionalities. To properly achieve the intended outcome, it is vitally necessary to focus one's attention solely on the persons who have passed the threshold of intrusion. This is a requirement that cannot be avoided. By establishing the direction of the movement, specifically, whether it is oriented in a downward or upward trajectory, it is possible to make the completion of this aim easier to execute. By utilizing bounding boxes, which enable a more comprehensive visual representation, it is possible to generate a visual depiction of the detection of human things. These bounding boxes intend to enclose the human items that have been recognized as belonging to these categories. Additionally, a one-of-a-kind identification number is attached to each bounding box. This number serves the purpose of indicating the order in which the human artifacts were found. The architecture of the system has been painstakingly designed by the specifications through the use of a cascade method. The methodology that has been proposed begins with the identification of a human entity, which is then followed by the monitoring and tracking of that individual throughout a series of frames that are located close to one another. A high level of proficiency in the execution of counting precisely and tracking a large number of items at the same time is demonstrated by the methodology that has been proposed. Additionally, it can efficiently cope with situations in which a human-made thing that has never been seen before comes into view, and it immediately assigns it a distinct identity. This is a capability that it possesses.

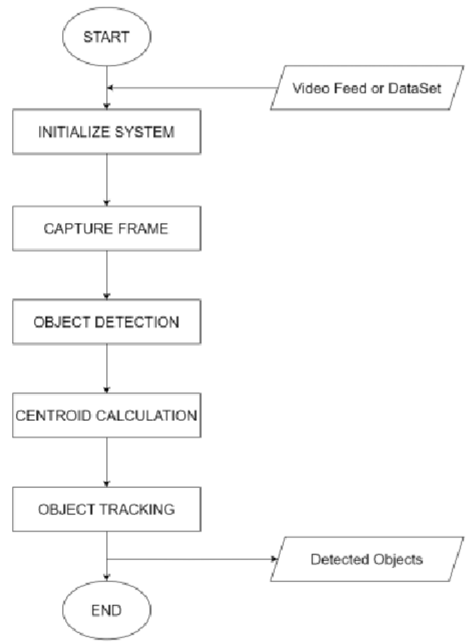


Fig. 1. Flow Chart

Therefore, to ensure the proper execution of this system, it is necessary to incorporate two essential components, which are as follows: both a human detection module that is only devoted to the identification of human beings and a tracking mechanism that is exclusively designed to monitor the human objects that have been recognized are included in this system.

IV. Methodology

By utilizing a computer vision investigation that is centered around the utilization of the Single Shot Detector (SSD) in conjunction with the Mobile Net architecture for real-time object detection , the researchers have used a computer vision investigation in the defining of the project. One of the most important aspects of improving the functionality of mobile devices, IP cameras, and scanners is the incorporation of this cutting-edge amalgamation, which efficiently optimizes both speed and efficiency demonstrated in Figure 3. In contrast to two-shot detectors like R-CNN, SSD can identify objects in a single iteration, which is a significant advantage over the latter. A sort of object detection model known as the SSD is distinguished from two-shot detectors by the strategy that it takes. In contrast to two-shot detectors, which

Algorithm 1 Object Tracking Algorithm

```
Initialize:
- List of detected objects
- Dictionary to store object IDs and their respective centroids
function DETECT_OBJECTS(frame)
    # Utilize Single Shot Detector (SSD) to detect objects
    detected_objects ← SSD.detect(frame)
    return detected_objects
end function
function TRACK_OBJECTS(detected_objects)
    for each object in detected_objects do
        if object is new then
            Assign a unique ID
            Add object ID and centroid to dictionary
        else
            Update centroid for existing object
        end if
    end for
end function
function DISPLAY_TRACKING_RESULTS(frame)
    for each object ID and centroid in dictionary do
        Draw bounding box around object using centroid
        Display object ID and bounding box on frame
    end for
end function
Main Loop:
while video is running do
    Capture frame from video
    detected_objects ← DETECT_OBJECTS(frame)
    TRACK_OBJECTS(detected_objects)
    DISPLAY_TRACKING_RESULTS(frame)
end while
```

Fig. 2. Algorithm

normally carry out the duties of region proposal and object identification in consecutive processes, SSD is capable of carrying out both functions concurrently. This indicates that SSD is capable of generating region proposals and detecting objects in a single run over the network, which ultimately results in processing that is both quicker and more precise. Real-time object detection is made possible in a variety of applications thanks to this approach, which eliminates the need for additional computing stages. As opposed to two-shot detectors like R-CNN, the efficiency of SSD is significantly higher. An important departure from the traditional multi-stage object detection methods is represented by the employment of SSD. By incorporating region proposal and object detection, which is made possible by the utilization of SSD, the system increases both its efficiency and its speed. The ability to make decisions quickly, based on the recognition of objects inside a given scene, is required for real-time applications for them to function well. Centroid tracker is a well-established method that is used to track objects across multiple frames. It is essential to make use of this tracking technique to maintain regularity and difference over consecutive frames to guarantee a thorough evaluation of the items selected by SSD. The determination of the centroid of a bounding box is considered to be the core idea that is at play here. It is possible to significantly improve the accuracy of object localization and tracking in visual images by employing this computational

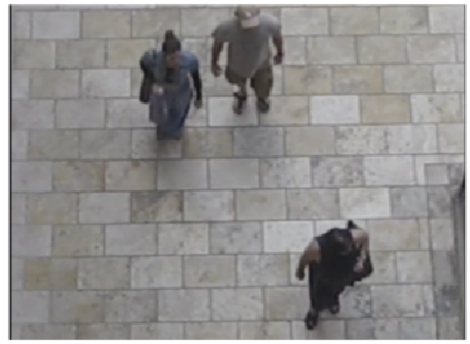


Fig. 3. Stock video

strategy. A centroid tracker is made use of to determine the centroid of the bounding boxes that are produced by the Single Shot MultiBox Detector (SSMD). To encapsulate and delineate items that have been successfully detected inside specific frames of a given dataset or video sequence, bounding boxes are comprised of a set of two-dimensional coordinates, which are denoted as (x, y). Bounding boxes serve the aim of encapsulating and outlining things. To find the centroid of the bounding boxes, the tracker technique makes use of a computational approach. This makes it possible to discover the point within the frame that is the most central to the item. By making use of the concept of an object's centroid, it is possible to achieve the capability of monitoring and determining the trajectory and spatial coordinates of an object across numerous frames.

To ensure that the Centroid Tracker can perform its functions correctly and effectively, every object that is being observed must be assigned a unique and customized identity demonstrated in Figure 4. This enables the system to monitor and record the object's spatial displacement over consecutive frames. The tracking mechanism of the system is dependent on the unique identification (ID) that is issued to each object. As a result of the system's consistent establishment of a correlation between the unique identification and the object in question, it can monitor and record the trajectory of the object throughout the procedure. In circumstances that involve several different items, the fact that it can independently track distinct things and provide a unique identity for each of them makes it extremely significant. The numerical IDs that have been provided are permanent, which ensures that continuity will be maintained and makes the process of reconstructing the line of travel of the object easier to accomplish.

V. Results Analysis

To accomplish the goals of this investigation, the system that is being studied is put through a series of tests on the graphics processing unit (GPU) of a computer that is located in the investigation area. The dataset that is being utilized in these experiments has been selected with considerable care. The central processing unit (CPU) was placed in the

Fig. 4. After applying algorithm

TABLE I
ACCURACY RESULT OF NORMAL MODEL WITH HIGHER ACCURACY

No.	Camera orientation	Expected up count	Expected down count	Actual up count	Actual down count
1	Front view	2	5	2	5
2	Front view	7	4	8	4
3	Overhead view	12	12	12	13
4	Overhead view	3	7	3	6
5	Overhead view	3	5	2	4

TABLE II
ACCURACY RESULT WITH FASTER PERFORMANCE

No.	Camera orientation	Expected up count	Expected down count	Actual up count	Actual down count
1	Front view	2	5	3	4
2	Front view	7	4	4	2
3	Overhead view	12	12	11	10
4	Overhead view	3	7	3	7
5	Overhead view	3	5	1	6

role of the target machine during the initial testing of the system. Regardless of this, it was quickly rejected because GPU displayed substantially higher levels of performance than previously thought. Only the number of line incursions going up and down are significant for counting. Tables I and II display the accurate manual count and the predicted count using the proposed technique for various camera angles. To get the final accuracy percentage, determine the error count by subtracting the expected count from the actual count. The final accuracy is 91.66% for the more precise normal model and 78.33% for the speedier model. Both calculations are detailed and presented individually below.

Error rate = sum of error count/(sum of expected up count and down count) × 100%
= (0+1+1+1+2)/(2+5+7+4+12+12+3+7+3+5)× 100%
= (5/60) × 10 %
= 0.0833 × 100%
= 8.33%
Overall accuracy = total expected count - error rate
= 100% - 8.33%
= 91.66%

Error rate = sum of error count/(sum of expected up count and down count)× 100%
= (2+5+3+0+3)/(2+5+7+4+12+12+3+7+3+5)× 100%
= (13/60)×100%
= 0.2166×100%
= 21.66%

Overall accuracy = total expected count - error rate
= 100% - 21.66%
= 78.33%

VI. CONCLUSION

The research embarked on a computer vision study aimed at leveraging the Single Shot Detector (SSD) approach alongside the Mobile Net architecture to enable real-time object detection. This integration significantly enhances speed and efficiency crucial for mobile devices, IP cameras, and scanners. Unlike two-shot detectors like R-CNN, SSD can identify objects in a single iteration, combining region proposal and object identification into a unified phase, eliminating the need for a separate region proposal network (RPN). This consolidation streamlines processing, optimizes resources, and simplifies the detection pipeline. SSD's efficacy surpasses that of traditional two-shot detectors like R-CNN, offering a notable departure from multi-stage object detection techniques. Its benefits extend to various computer vision tasks, enhancing both efficiency and speed, crucial for real-time applications requiring swift decision-making based on object recognition within a scene.

The proposed solution incorporates a centroid tracker for object tracking across successive frames, ensuring consistency and clarity in evaluating items identified by the Single Shot Detector (SSD). Centroid computation within bounding boxes enhances tracking accuracy and geographic localization of objects. By utilizing a centroid tracker alongside the SSD, the system computes centroids to encapsulate and track identifiable objects across frames. Each object is assigned a unique identification (ID) by the tracker, facilitating independent tracking and consistent recording of object movement over time. This unique ID system enables efficient tracking of multiple objects by maintaining a constant relationship between the ID and the object, ensuring uninterrupted record-keeping and simplifying the process of recreating object trajectories. The permanent nature of these identification numbers streamlines the tracking process, making it easier to reconstruct object movements over time.

REFERENCES

[1] Chun, Sung, and Chan-Su Lee. "Applications of human motion tracking: Smart lighting control." Proceedings of the IEEE conference on computer vision and pattern recognition workshops. 2013. Based on Multiple Aspects Using Image Processing Techniques", IEEE Access, 2022.

[2] Liu, J., Liu, Y., Cui, Y., and Chen, Y. Q. (2013, September). Real-time human detection and tracking in complex environments using single RGBD camera. In 2013 IEEE International Conference on Image Processing (pp. 3088-3092). IEEE.

[3] Elsisi, Mahmoud, Minh-Quang Tran, Karar Mahmoud, Matti Lehtonen, and Mohamed MF Darwish. "Deep learning-based industry 4.0 and internet of things towards effective energy management for smart buildings." Sensors 21, no. 4 (2021): 1038.

[4] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.

[5] Ahmadi-Karvigh, Simin, Ali Ghahramani, Burcin Becerik-Gerber, and Lucio Soibelman. "Real-time activity recognition for energy efficiency in buildings." Applied energy 211 (2018): 146-160.

[6] Liu, Wei, et al. "Ssd: Single shot multi-box detector." Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016.

[7] Lu, S., Hameen, C.E. and Aziz, A., 2018, January. Dynamic hvac operations with real-time vision-based occupant recognition system. In 2018 ASHRAE Winter Conference, Chicago.

[8] Abd-Almageed, Wael, Mohamed Hussein, Mohamed Abdelkader, and Larry Davis. "Real-time human detection and tracking from mobile vehicles." In 2007 IEEE Intelligent Transportation Systems Conference, pp. 149-154. IEEE, 2007.

[9] Viraktamath, S. V., Mukund Katti, Aditya Khatawkar, and Pavan Kulkarni. "Face detection and tracking using OpenCV." The SIJ Transactions on Computer Networks and Communication Engineering (CNCE) 1, no. 3 (2013): 45-50.

[10] He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (pp. 2961-2969).

[11] Pennisi, Andrea, Domenico D. Bloisi, and Luca Iocchi. "Online real-time crowd behavior detection in video sequences." Computer Vision and Image Understanding 144 (2016): 166-176.

[12] Mokayed, Hamam, Tee Zhen Quan, Lama Alkhaled, and V. Sivakumar. "Real-time human detection and counting system using deep learning computer vision techniques." In Artificial Intelligence and Applications, vol. 1, no. 4, pp. 221-229. 2023.

[13] Girshick, R. (2015). Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision (pp. 1440-1448).

[14] Arroyo, Roberto, J. Javier Yebes, Luis M. Bergasa, Iván G. Daza, and Javier Almazán. "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls." Expert systems with Applications 42, no. 21 (2015): 7991-8005.

[15] APAO, NORMA J., and VICTOR JOHN L. ANUNCIADO. "Intelligent Energy Saving System using Real-Time Moving Object Detection and Microcontroller Unit." ACADEME 11 (2017).

[16] Chandrasiri, Amila Prasad, and Devindi Geekiyanage. "Real-time object detection system for building energy conservation: an IP camera based system." THIRTY-FOURTH ANNUAL CONFERENCE. 2018.

[17] Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., and Belongie, S. (2017). Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2117-2125).

[18] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems 25 (2012).

Human Count SK

ORIGINALITY REPORT

8%	7%	3%	1%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	ojs.bonviewpress.com Internet Source	3%
2	documents.mx Internet Source	1%
3	www.interscience.in Internet Source	1%
4	academic.oup.com Internet Source	<1%
5	www.glbitm.org Internet Source	<1%
6	authors.library.caltech.edu Internet Source	<1%
7	Submitted to CSU, San Jose State University Student Paper	<1%
8	www.readkong.com Internet Source	<1%
9	dokumen.pub Internet Source	<1%

10	mafiadoc.com Internet Source	<1%
11	cvpr.thecvf.com Internet Source	<1%
12	www.researchgate.net Internet Source	<1%
13	"Machine Learning and Cognitive Computing for Mobile Communications and Wireless Networks", Wiley, 2020 Publication	<1%

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	On		