**Literature Review**

Researchers have explored the use of vehicle trajectory data in traffic conditions, utilizing unmanned aerial vehicles (UAV) and videos for analysis[1]. They proposed a vehicle trajectory model incorporating CNN and YOLO v3 for vehicle detection, yielding accurate results. Stahl et al. addressed CNN limitations by designing a YOLOv3-based model for fast detection on the PASCAL VOC dataset, focusing on reducing power consumption[1].

A new method for real-time crowd detection from videos was introduced, emphasizing the importance of physical distancing in indoor spaces, especially during events like the COVID-19 pandemic.[2]

Machine learning and computer vision are crucial in data sensing, understanding, and action-taking based on past outcomes. Various methods like supervised learning, unsupervised learning, semi-supervised learning, and machine learning algorithms are employed for prediction and analysis. Object detection, traffic detection models, and cloud-based machine learning services are highlighted in the research[3].

Computer vision and machine learning are crucial areas of research that have seen significant advancements in recent years. Computer vision involves using images and pattern mappings to automate monitoring, inspection, and surveillance tasks. It treats images as arrays of pixels and has led to the automatic analysis and annotation of videos, showcasing capabilities like classification, object detection, and instance segmentation. The integration of machine learning paradigms such as support vector machines, neural networks, and probabilistic graphical models has further enhanced the capabilities of computer vision systems. The evolution of machine learning has transformed computer applications from simple data processing to more sophisticated systems that can learn and improve over time. Western countries have shown a keen interest in machine learning, computer vision, and pattern recognition, which is evident through various conferences, workshops, and real-life implementation. Recent applications of machine learning in computer vision include object detection, object classification, and extracting relevant information from images and videos, showcasing the practical implications of these technologies[4]

Additionally, tools like Tensorflow, Faster-RCNN-Inception-V2 model, and Anaconda have been utilized to identify objects like cars and persons in images, highlighting the practical implementations of machine learning in computer vision.

Machine learning techniques like the nearest neighbour algorithm, support vector machine (SVM), decision tree (DT), random forest, and Naïve Bayes classifier have been previously employed for land cover prediction. Input features are typically gathered from satellite images using time-series data after normalized difference vegetation index, with output classes including impervious, forest, grass, water, orchard, and farm. Synthetic minority techniques are applied for data balancing using oversampling, with k-NN providing the highest accuracy. High spatial resolution images from Google Earth create land cover thematic maps for urban scenes. Different image classification methods are used to classify land usage and cover maps, with maximum likelihood techniques demonstrating better accuracy and kappa coefficient results. Using minimum distance and support vector machines as supervised classifiers, along with maximum likelihood and parallelepiped systems, has shown improved kappa coefficient and overall accuracy results.[5]

In the context of the COVID-19 pandemic, maintaining physical distance is crucial, leading to the need for limited people in indoor spaces, as recommended by the World Health Organization. The capacity of indoor areas varies based on their size, necessitating the measurement of indoor spaces to calculate the maximum number of people allowed. Utilizing computers and cameras can aid in enforcing capacity rules in indoor spaces by measuring specific regions in real-time and counting the number of people present. The proposed method involves predetermining the borders of a region in a video, identifying and counting people within that region, estimating its size, and determining the maximum number of people it can accommodate. The study employed the YOLO object detection model, specifically YOLO v3 and YOLO v5s, using Microsoft COCO dataset pre-trained weights to detect and label persons in the specified region, with performance metrics such as mean average precision (mAP), frame per second (fps), and accuracy rate is evaluated. The YOLO v3 model demonstrated the highest accuracy rate and mAP metrics. At the same time, YOLO v5s excelled in fps rate among non-Tiny models, showcasing the effectiveness of these models in real-time crowd detection from videos.[1], [2]

Object detection is crucial for real-world applications such as surveillance, security, and automated vehicle systems. The paper compares the effectiveness of two algorithms, YOLOv3 and SSD, for people detection and counting at junctions. The study involves two main tasks: object detection using image datasets and counting objects using video datasets. Results show that SSD outperforms YOLOv3 in terms of precision, recall, and F1 measure for people detection and counting. Accurate people detection and counting are essential for applications like surveillance and security systems, and people counting at junctions serves various purposes in ensuring task integrity. The research provides a detailed analysis of the efficiency of YOLOv3 and SSD models for detection and counting tasks. The paper contributes by evaluating and comparing YOLOv3 and SSD models, offering insights for future research in the field of people detection and counting.[1]

Computer vision aims to identify and recognize objects of various characteristics such as size, shape, and position. One of the challenges faced in computer vision is dealing with issues like illumination and viewpoint of objects. Deep learning techniques, particularly Convolutional Neural Networks, have shown high accuracy and precision in detecting and recognizing objects under such conditions. The proposed work in this paper focuses on object detection in a college environment, involving identifying individuals wearing ID cards using TensorFlow object detection API and recognizing faces using the Haar cascade method of OpenCV. Deep learning, explicitly using Convolutional Neural Networks, has been widely adopted in various computer vision tasks due to its effectiveness in handling complex visual data and achieving high levels of accuracy and precision. The combination of TensorFlow and OpenCV in this study showcases the integration of deep learning and traditional computer vision techniques for efficient object detection and recognition tasks in real-world scenarios.[4]

Security is one of the significant concerns of humans, as well as automated surveillance and security in improving home-based safety through the development of various safety schemes based on technologies such as embedded Linux, Raspberry Pi, OpenCV, and face recognition. It discusses the use of face recognition technology to develop a safety access control application, utilizing algorithms like Haar-like features, HOG, and SVM for face detection and recognition. The research paper suggests a hybrid expression recognition algorithm incorporating PCA and LDA for improved system control and verification performance. Emphasis is placed on the importance of developing efficient and accurate expression recognition schemes for safety controllers, especially in regions like Nigeria, to combat criminal activities. The paper also explores the use of low-cost, real-time structural designs based on FPGA for expression recognition, which is beneficial for identifying faces in various scenarios. It highlights the significance of advancements in face recognition technology for enhancing safety measures and providing flexibility in smart home control systems, mainly through Raspberry Pi boards. Additionally, the research delves into the application of fog computing for expression recognition and resolution outline to assess the impact of safety schemes on system performance.[6]

|  |  |  |  |
| --- | --- | --- | --- |
| **Year of Publication** | **Author / Company** | **Techniques/ Algorithms** | **Gaps/drawbacks** |
| 2017 | Machine Learning in Computer Vision, Asharul Islam Khan, Salim Al-Habsib | Techniques: supervised, unsupervised, semi-supervised.  Algorithms: neural networks, k-means clustering, support vector machine. | Lack of focus on real-time applications in machine learning.  Limited discussion on the challenges of data annotation in computer vision. |
| 2019 | Face Detection and Recognition for use in  Campus Surveillance, B. Kranthikiran, Padmaja Pulicherla | Techniques: Geometric, feature-based, template matching.  Algorithms: PCA, ICA, LDA, SVM, deep neural networks. | Lack of discussion on algorithm performance evaluation and comparison.  Absence of exploration on scalability and adaptability of the system. |
| 2020 | Computer Vision Based Campus Surveillance,  Mahesh Mahajan,Lokesh Sharma | Machine learning techniques: SVM, KNN, Naive Bayes, Decision Tree. | Lack of comparison with existing surveillance systems |
| 2020 | People detection and counting using YOLOv3 and SSD models | YOLOv3 and SSD are object detection algorithms used in the research.  Object detection and counting are tasks performed by the algorithms. | Lack of exploration of real-world applications beyond surveillance and security.  Limited discussion on the impact of dataset variations on model performance. |
| 2020 | ID Card Detection with Facial Recognition using  Tensorflow and OpenCV  Kushal M Kushal Kumar B V Charan Kumar M J Prof.Pappa M | Object detection using tensor flow for ID card recognition.  Haar cascade method in OpenCV for face detection.  Mobile-based face recognition system using Euclidean distance calculation. | Lack of discussion on real-world implementation challenges.  Absence of comparison with existing ID card detection methods |
| 2023 | A new YOLO‑based method for real‑time crowd detection from video and performance analysis of YOLO models  Mehmet Şirin Gündüz· Gültekin Işık | YOLO models were used for real-time crowd detection in the study.  Deep learning algorithms applied in various fields including image processing. | Lack of comparison with other object detection models.  Absence of discussion on the impact of varying video resolutions. |

**References**

[1] P. Gupta, V. Sharma, and S. Varma, “People detection and counting using YOLOv3 and SSD models,” *Mater Today Proc*, 2021, doi: 10.1016/j.matpr.2020.11.562.

[2] M. Ş. Gündüz and G. Işık, “A new YOLO-based method for real-time crowd detection from video and performance analysis of YOLO models,” *J Real Time Image Process*, vol. 20, no. 1, Feb. 2023, doi: 10.1007/s11554-023-01276-w.

[3] A. I. Khan and S. Al-Habsi, “Machine Learning in Computer Vision,” in *Procedia Computer Science*, Elsevier B.V., 2020, pp. 1444–1451. doi: 10.1016/j.procs.2020.03.355.

[4] Institute of Electrical and Electronics Engineers and RVS College of Engineering & Technology, *Proceedings of the 2nd International Conference on Inventive Research in Computing Applications (ICIRCA 2020) : 15-17 July, 2020*.

[5] M. Mahajan, R. Rajeshwaribajre, L. Sharma, and S. L. Varma, “Computer Vision Based Campus Surveillance,” in *Proceedings - IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2020*, Institute of Electrical and Electronics Engineers Inc., Dec. 2020, pp. 1013–1018. doi: 10.1109/ICACCCN51052.2020.9362776.

[6] P. Pulicherla, “Face Detection and Recognition for use in Campus Surveillance International Journal of Innovative Technology and Exploring Engineering (IJITEE),” *Article in International Journal of Innovative Technology and Exploring Engineering*, 2020, doi: 10.35940/ijitee.B7682.019320.