Department of Electrical, Computer, and Software Engineering

Part IV Research Project

Literature Review and

Statement of Research Intent

Project Number: 73

[Title of Report]

Adwait Mane

Project Partner: Vinayak Joshi

Supervisor: Robert Amor

22/04/2024

Declaration of Originality

This report is my own unaided work and was not copied from nor written in collaboration with any other person.

[Signature]

Name: Adwait Mane

1. Introduction

Electronic scooters (e-scooters) have rapidly become a common sight in major cities across the world. Their convenience, cost-effectiveness, environmental benefits, and sheer fun factor have fuelled an exponential increase in e-scooter riders. Although e-scooters can be bought privately, they have been made incredibly accessible through rideshare apps like Uber and Lime. In 2023, Lime set a record 156 million e-scooter trips, about 5 trips every second for the entire year [1].

The widespread adoption of e-scooters as a micro-mobility solution reflects a growing desire within cities to address traffic congestion, minimise emissions, and improve the efficiency of urban transportation. Their ease of use and proliferation of short-term rental services contribute to studies suggesting that e-scooters may replace passenger cars for short-distance trips less than two kilometres [2].

However, the recent surge in e-scooter popularity has brough both convenience and new challenges to urban environments. While offering a quick and eco-friendly mode of transportation, e-scooters can also create unique security concerns. By exploiting the agility of an e-scooter, there have been incidents of intruders using e-scooters to gain quick access to restricted areas, such as parking garages, and engaging in theft or vandalism before traditional security measures can respond. The speed and size of e-scooters can make them difficult to track and monitor, creating a need for automated security solutions that utilize surveillance cameras to efficiently detect and identify potential threats.

Object detection, a field within artificial intelligence (AI), offers a promising approach to address these security challenges. By developing algorithms capable of accurately identifying e-scooters and their riders in real-time, object detection can trigger timely alerts within security monitoring systems. This technology has the potential to enhance security in a variety of settings, from parking garages to public spaces, and contribute to a safer urban environment.

2. Research Objectives

This research aims to advance the field of object detection by developing and optimizing techniques specifically tailored to the accurate and efficient identification of e-scooters and their riders in security camera feeds. The project has two primary components with interlinked outcomes:

Thorough research will delve into the evolution of object detection methodologies, current real-time techniques, and state-of-the-art algorithms. The review will critically analyse strengths, weaknesses, potential gaps, and performance metrics, focusing on their suitability for the specific challenges of e-scooter detection. Additionally, the review will emphasise integration challenges faced by real-world applications of object detection systems (cost, deployment complexity, resource constraints). A core focus will be the critical assessment of standard performance metrics, questioning their relevance and potential biases in contexts demanding real-time accuracy. This review aims to establish a theoretical framework to guide the development of a tailored solution.

Building upon insights from the literature review, the research will explore the potential for optimising existing detection techniques or developing novel approaches. The focus will be on addressing the need for balanced performance metrics that ensure both accuracy and real-time applicability in a security context.

3. Literature Review

3.1. Object Detection Definition and Importance

Object detection is a foundational part of computer vision, allowing machines to understand and interpret visual information. Its main objective is to identify and pinpoint specific objects of interest within images and video frames [3]. At its core, object detection asks two fundamental questions:

Where? With precise accuracy, object detection systems locate objects and outline their boundaries using bounding boxes. These boxes precisely mark the extent of the object within the image.

What? Beyond localisation, object detection assigns a class level to each identified object. This classification allows the system to not only recognise the presence of an object, but also understand its nature. For instance, the system might classify a detected object as a “person”, “car, “scooter”, or other predefined categories.

Object detection serves important for advancements across diverse fields. For example, in medical scans, object detection can assist in pinpointing tumours, organs, or other areas of interest, aiding in diagnosis and treatment planning. In retail, object detection can be used to automate checkout systems and inventory management.

One field that can significantly benefit from object detection is security and surveillance. Real-time object detection is crucial for identifying potential threats, monitoring restricted areas, and tracking suspicious activities. Traditional security systems solely reliant on human monitoring are prone to errors due to fatigue, limited attention span, and other factors. Object detection can significantly reduce false alarms. When integrated with surveillance footage, it can be used to identify and tag objects of concern and raise alarms only when relevant objects are confirmed [4]. By integrating object detection with cameras, alarms, access control and communication tools, existing systems become a comprehensive security network. This allows for automated responses, swift alerts, and a robust approach to safeguarding people and assets.

3.2. Traditional Object Detection Methods

Before deep learning approaches, traditional object detection relied on handcrafted features and specialised algorithms [5]. Some notable techniques include:

Histogram of Oriented Gradients (HOG): HOG captures gradient orientation within localised portions of an image, making it useful for detecting edges and boundaries. This method was widely used for pedestrian detection [6].

Haar-like Features: These rectangular filters compute intensity differences within regions. Used in the Viola-Jones algorithm, they enabled fast face detection [7].

Local Binary Patterns (LBP): A texture descriptor that assigns a binary code to local image regions, historically used in object detection for its efficiency in capturing texture and illumination invariance [8].

While these traditional methods offered some success in specific scenarios, their reliance on handcrafted features often limited their accuracy, flexibility, and ability to handle complex visual environments.

3.3. Modern Object Detection Methods

Object detection has witnessed a dramatic transformation, powered by deep learning methods. Convolutional Neural Networks (CNNs) have an inherent ability to mimic the hierarchical nature of human visual processing, allowing them to automatically learn intricate features from image data [9]. Combined with an abundance of labelled datasets and updates in device hardware, deep learning object detection algorithms have become remarkably accurate and efficient, offering competitive real-time detection performance compared to traditional methods.

R-CNN: This algorithm pioneered the use of CNNs for object detection, using a selective search algorithm to generate region proposals, which are then independently classified by a CNN. It has improved detection accuracy over traditional methods due to CNN-based feature extraction [9]. However, its multi-stage process (region proposal, classification, regression) made it computationally expensive and slow for real-time use.

Fast R-CNN: This algorithm streamlines the process by first computing a feature map for the entire image then mapping region proposals onto this map. A RoI pooling layer extracts features for each region, followed by classification and regression stages [9]. This was significantly faster than R-CNN by sharing computation and enabling end to end training. However, it still relies on a separate region proposal step, which becomes a computational bottleneck.

Faster R-CNN: This eliminates the region proposal bottleneck by introducing a Region Proposal Network (RPN). The RPN is a fully convolutional network that directly generates region proposals from the feature map. It offers a unified, end-to-end trainable model for both region proposal and object detection, drastically improving speed and accuracy [10]. Though it is faster than its previous iterations, it is a two-stage process, having some computational overhead compared to single-stage detectors.

YOLO (You Only Look Once): YOLO uses a single neural network and regression technique, directly outputting the bounding boxes and class probabilities in a single evaluation. It divides the input image into a grid of cells, where each grid cell is responsible for predicting bounding boxes (coordinates and dimensions), class probabilities for the detected object, and confidence scored reflecting the likelihood of an object being present [11]. The single-shot architecture is fast, making YOLO ideal for real-time applications. However, early YOLO versions struggled with accurately detecting small, especially grouped objects. This is due to the grid-based approach, limiting the number of objects each cell can predict, impacting detection in crowded scenes.

YOLOv3: This algorithm builds upon the core concept of YOLO as a single-stage, regression-based object detector. It introduced significant advancements to enhance accuracy, particularly in detecting both small and large objects. YOLOv3 makes predictions at three different scales, allowing it to effectively detect objects of varying sizes. It uses Darknet-53, a deeper CNN with residual connections as its feature extraction backbone. This enables the model to learn richer and more complex representations of the image data. YOLOv3 maintains the real-time speed advantage from its previous version, with improved accuracy and the ability to handle scenarios where objects might have multiple associated labels [13, 14]. However, it may still have a slight accuracy disadvantage compared to two stage detectors like Faster R-CNN in some complex scenarios.

While these methods represent widely used approaches, the field of object detection is constantly evolving. Newer YOLO iterations (up to YOLOv8) and other architectures like RetinaNet offer promising, yet largely untested potential for e-scooter detection. RetinaNet's success in detecting e-scooters in [15] highlight the value of exploring diverse algorithms specifically tailored to the challenges of e-scooter detection within varied environments.

3.4. Current Literature on E-Scooter Detection

3.5. Performance Metrics

3.6. Common Findings

3.7. Contradictive Findings

3.8. Datasets

3.9. Suitability within Security Context

[1]Lime, “We Achieved A Record Setting Year In 2023, With Highest Ever Total Rides And Gross Bookings,” Apr. 08, 2024. https://www.li.me/blog/lime-achieves-record-setting-year-in-2023-with-highest-ever-total-rides-and-gross-bookings-as-it-continues-to-set-the-pace-for-shared-electric-vehicle-industry

[2]C. S. Kopplin, B. M. Brand, and Y. Reichenberger, “Consumer acceptance of shared e-scooters for urban and short-distance mobility,” Transportation Research Part D: Transport and Environment, vol. 91, p. 102680, Feb. 2021, doi: https://doi.org/10.1016/j.trd.2020.102680.

[3]J. Kaur and W. Singh, “Tools, techniques, datasets and application areas for object detection in an image: a review,” Multimedia Tools and Applications, Apr. 2022, doi: https://doi.org/10.1007/s11042-022-13153-y.

[4] Algotive AI , “Object Recognition in Security: Everything You Need to Know,” www.algotive.ai, Sep. 07, 2023. https://www.algotive.ai/blog/object-recognition-in-security

[5]N. Alsharabi, “Real-Time Object Detection Overview: Advancements, Challenges, and Applications,” JOURNAL OF AMRAN UNIVERSITY J. Amr. Uni, vol. 03, p. 267, 2023.

[6]N. Dalal and B. Triggs, “Histograms of Oriented Gradients for Human Detection,” 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), vol. 1, pp. 886–893, 2005, doi: https://doi.org/10.1109/cvpr.2005.177.

[7]S. O. Adeshina, H. Ibrahim, S. S. Teoh, and S. C. Hoo, “Custom Face Classification Model for Classroom Using Haar-Like and LBP Features with Their Performance Comparisons,” Electronics, vol. 10, no. 2, p. 102, Jan. 2021, doi: https://doi.org/10.3390/electronics10020102.

[8]G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu, “Boosting Local Binary Pattern (LBP)-Based Face Recognition,” Advances in Biometric Person Authentication, pp. 179–186, 2004, doi: https://doi.org/10.1007/978-3-540-30548-4_21.

[9] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: https://doi.org/10.1038/nature14539.

[10] R. Girshick, “Fast R-CNN,” 2015 IEEE International Conference on Computer Vision (ICCV), pp. 1440–1448, Dec. 2015, doi: https://doi.org/10.1109/iccv.2015.169.

[11]S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” arXiv.org, 2015. https://arxiv.org/abs/1506.01497

[12]J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 779–788, 2016, doi: https://doi.org/10.1109/cvpr.2016.91.

[13]P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, “A Review of Yolo Algorithm Developments,” Procedia Computer Science, vol. 199, pp. 1066–1073, Jan. 2022, doi: https://doi.org/10.1016/j.procs.2022.01.135.

[14] H. H. Nguyen, M. H. Nguyen, and Q. Sun, "Electric Scooter and Its Rider Detection Framework Based on Deep Learning for Supporting Scooter-Related Injury Emergency Services," in Geometry and Vision, 2021, pp. 233-246.

[15] Cal Poly Digital Transformation Hub, “E-Scooter Counting on Sidewalks – DxHub,” Cal Poly Digital Transformation Hub. https://dxhub.calpoly.edu/challenges/escooter-counting-on-sidewalks/ (accessed Apr. 24, 2024).