Challenging HOG's Pedestrian Detection Accuracy

Sherdan Caruana
Institute of Information & Communication Technology
Malta College or Arts, Science & Technology
Corradino Hill
Paola PLA 9032
sherdan.caruana.d57125@mcast.edu.mt

***Abstract*—This project aims to tackle the issue of pedestrian safety in urban environments, where elements such as high traffic volumes, complex road networks, and varying lighting conditions significantly contribute to numerous accidents. This research proposes to explore and experiment on the development of pedestrian detection using Python as the primary programming language, while implementing OpenCV and the Histogram of Oriented Gradients (HOG) for feature extraction and detection. By using OpenCV and the HOG algorithm, the study aims to create a pedestrian detection system capable of accurately identifying pedestrians in diverse environments and conditions. In addition to algorithm development, this research conducted will include testing under varying conditions, including different lighting scenarios and video resolutions, to evaluate the robustness and effectiveness of the pedestrian detection system. These experiments are aimed to test the system's performance and identify areas for improvement, ultimately contributing to the advancement of pedestrian safety technology in urban environments. The research faced challenges with false positives throughout the testing phase due to the multiple varying challenges it was tested against to identify the algorithm's flaws.**

***Index Terms*—Pedestrian detector, HOG, Computer vision**

I. Introduction

Road safety is a massive concern in urban environments such as cites and any densely populated areas with access to roads networks. Given the multiple challenges encountered when dealing with such a problem such as the large number of vehicles on the road along with the density of pedestrians in large, populated areas where typically accidents are most likely to happen. To address these challenges this project aims to develop a well-founded pedestrian detection system with the use of computer vision technology. While sticking to this goal the project will be created and adapted while utilizing the Histogram of Oriented Gradients (HOG) algorithm. By using Python with the OpenCV library, the project aims to create a well-rounded detection system capable of accurately identifying pedestrians in real-time video streams. This research will explore various approaches to creating and experimenting with detection systems, ultimately arriving at a conclusion regarding the most effective method among them. Experiments to find out how such detection systems will perform will be conducted, such as assessing the impact of variating lighting conditions on the detection process during both daytime and nighttime scenarios. Additionally, tests will be carried out to determine how compression and resolution affect the quality and effectiveness of the detection process. Furthermore, comparisons will be made between lower-quality videos and higher-quality ones to find out their respective effects on the detection process.

II. Literature Review

Pedestrian detection stands as a foundational challenge for researchers aiming to study deep learning algorithms and develop safer forms of travel. Zhang et al. introduced CityPersons as a new set of high-quality annotations aimed to provide a dataset to adapt the FasterRCNN architecture to achieve cutting-edge results on established benchmarks like Caltech and KITTI [1]. The CityPersons dataset was created to provide high-quality bounding box annotations for pedestrians in urban environments, aiming to enhance pedestrian detection research by offering diverse and challenging training data for improved model performance and generalization across multiple benchmarks. The study demonstrated that by adapting the FasterRCNN architecture and training it on the CityPersons dataset, significant improvements were achieved in pedestrian detection performance. The adapted FasterRCNN model showcased state-of-the-art results on established benchmarks like Caltech and KITTI, particularly excelling in detecting small-scale pedestrians and handling heavy occlusions. The FasterRCNN detector, known for its competitive performance in general object detection had underperformed in pedestrian detection tasks on the Caltech dataset due to its inability to detect small-scale objects (50 70 pixels), which are prevalent in the dataset. The adapted FasterRCNN model demonstrated superior performance in pedestrian detection tasks, achieving a mean MR of 10.27 on the Caltech dataset and 12.81 on the CityPersons dataset [1]. These results indicate that the model excelled in detecting pedestrians in urban environments, with a slightly higher detection accuracy on the Caltech benchmark compared to the CityPersons dataset.

In recent studies on pedestrian detection, Liu et al. has introduced a technique known as Center and Scale Prediction (CSP). Liu et al managed to advance the current state-of-the-art in pedestrian detection by proposing a new perspective in high-level semantic feature detection CSP. The authors undertake experiments on two of the most popular pedestrian datasets, the Caltech and CityPersons datasets. The Caltech dataset contains around 2.5 hours of footage taken from autonomous driving, which is extensively labelled [2]. The CityPersons dataset is a subset of the Cityscapes dataset and

contains 2975 training images, 500 validation images and 1575 testing images, with an average of 7 pedestrians in each. The authors chose these datasets since they provide bounding boxes which align well with the centres of pedestrians. The proposed solution comprises of two components, the feature extraction, and the detection head. The feature extraction modules concatenate feature maps of different resolutions into a single map which is then fed to a 3x3 convolutional layer in the detection head, followed by two prediction layers. The prediction layers focus on the centre location and the corresponding scale. The output is a centre and heat map. The method was implemented in Keras using ResNet and ImageNet as the backbone architectures. The Adam network optimizer was selected. Training for CalTech was done in a mini-batch of 16 images on a GTX 1080Ti GPU with a learning rate of 0.001, stopping at 15,000 iterations. Training for CityPersons was done using a learning rate of 0.0002 with 2 images per GPU, utilising 4 GPUs and stopping at 37,500 iterations. The proposed system advanced the current state-of-the-art by reducing the miss rate on the Caltech dataset to 4.5% from 5.0% achieved by RepLoss. Experiments on CityPersons reduced the miss rate to 3.8% from 4% of RepLoss and 4.1% of OR-CNN. The authors have published their work with code on a Git repository which has also been integrated into the Pedestron (https://github.com/hasanirtiza/Pedestron) project.

Zhang et al. conducted a study on pedestrian detection using deep learning techniques, while also bringing up critical role's computer vision can take up in today's world to enhance safety. However, it introduces new deep learning techniques with the approach of CSANet in the aim of incorporating dual attention mechanisms to enhance the representation of feature maps. Another particular aim of this paper is to improve on AdaptFasterR-CNN which uses the CityPersons dataset to train the model, which was designed to utilize the dataset for strong generation capabilities. CSANet's proposed method is targeted to be anchorfree, being unrestricted by a predefined anchor box ratio and instead attempts to predict bounding boxes and key points to make up objects [3]. A study was conducted in aim of evaluating CSANet using the CityPersons dataset, analysing its components like feature extraction, fusion, and attention modules. Results show that in the experiment a NVIDIA GTX 1080Ti was used with 2 mini batches of images were used from the CityPersons dataset which resulted in a unique test time of 270ms, meaning CSANET came second to the CSP model that was trained with batch 8. Integration of CAM (Channel Attention Module) and SAM (Spatial Attention Module) into the ResNet-50 backbone of CSANet which serves as an enhancement feature in the aim of improving the performance of the model by effectively enhancing high-level semantic features and the ability to capture long-range dependencies with feature maps. With the integration of CAM and SAM into the backbone of CSANet, the model can effectively capture both channel-wise and spatial attention, leading to improved pedestrian detection performance [3].

A common challenge researchers face in detection is accurately maintaining bounding boxes around the intended target. Liu et al. sought to address the challenge of pedestrian detection in crowded scenes by introducing the Adaptive-NMS algorithm. This algorithm refines bounding boxes based on target density. An experiment was conducted to validate the effectiveness of the Adaptive-NMS algorithm in enhancing pedestrian detection performance in crowded scenes using different datasets which can affect its deep learning algorithms to detect target pedestrians in a big crowd of people. They had used 2 datasets for their evaluations, CityPersons dataset and the CrowdHuman dataset. The CityPersons dataset had contained 5,000 images with annotations for approximately 35,000 labelled persons and 13,000 ignored region annotations while the CrowdHuman Dataset contained 15, 000,4,370 and 5,000 images that are found trough out the internet. The CrowdHuman dataset contained images with much denser populations compared to those in CityPersons. However, CityPersons encompassed a broader range of weather conditions across 18 cities in Germany with an average of 7 pedestrians in average per image. With a base learning rate set to and 0.002 for FPN and RFB and with 4 Titan X GPUs the results had shown that using this method with CrowdHuman dataset should return a higher detection rate for targets in large crowds [4].

A study conducted in 2010 by Pang et al. dives into the histogram of oriented gradients (HOG) and Support Vector Machine (SVM) and how they could be improved to detect people effectively The paper proposes two methods to resolve the time-consuming detection process and attempt to speed up the process. The first method involves reusing features in blocks to construct HOG features for intersecting detection windows, while the second method utilizes sub-cell-based interpolation to efficiently compute HOG features for each block. These methods were tested using the INRIAdataset which contains around 2,308 images of pedestrians captured from different parts of a car allowing for multiple angled images to be used and tested with [5]. In their used cameras and sensors to capture different-resolution images of pedestrians for their methods to be experimented on. The results show that their proposed methods had yielded more detection results for the amount of performance that had previously been given without their methods being utilized. Combining the two ways results in more than five times increase in detecting humans in a 320 240 image.[6] Both methods combined the reduction of computational cost while increasing the detection accuracy in many cases.

So far, numerous datasets that have been created to tackle pedestrian detection across diverse weather conditions, varying light intensities, and complex environmental contexts. A standout dataset, the EuroCity Persons dataset, distinguishes itself in its specialization for traffic scenarios. Braun et al. had attempted to design the EuroCity Persons dataset to stand out as an invaluable resource for training models tailored to the complexities of detecting pedestrians amidst vehicular traffic [6]. The EuroCity Persons dataset comprises over 238,200 manually labelled person instances in over 47,300

images, collected from vehicles across 31 European cities. Using this as a benchmark to train models to detect pedestrians they had experimented on Faster R-CNN by integrating the region proposal network (RPN) into Faster R-CNN, with aim of increased efficiency. They had utilized region proposal network (RPN) to generate region proposals directly from the convolutional feature maps. Experiments conducted with EuroCity Persons dataset demonstrated the effectiveness of the optimized Faster R-CNN in traffic scenarios and complex environments. To conduct such experiments an Intel(R) Core(TM) i7-5960X CPU with a GTX TITANX 12GB was utilized with the possibility of Improving the runtime by replacing the convolutional neural networks (CNN) VGG base architecture with the GoogLeNet model. The results had shown that the Faster R-CNN model achieved a log-average-miss-rate of 7.9, 17.0, and 33.2 on the ”reasonable,” ”small,” and ”occluded” test cases, respectively [6]. Despite having more training data in the EuroCity Persons dataset compared to other datasets with the deep learning methods not being able to fully utilizes the potential in detection leaving room for improvement.

III. Research Methodology

The development of pedestrian detectors serves a critical role in safety especially when it comes to urban areas with multiple pedestrians with crowded sidewalks and intersections. In these environments, the development of such detection systems aids multiple drivers in avoiding accidents with high traffic volumes, complex road networks and constant danger of pedestrians crossing streets. Testing pedestrian detection systems in multiple scenarios is important in ensuring their reliability and effectiveness across multiple environments and conditions The prototype script is written in Python and consists of using OpenCV, Histogram of Oriented Gradients (HOG) to detect pedestrians within a specified video file. OpenCV provides comprehensive support for computer vision, offering functionalities and algorithms for image and video processing. HOG is used to capture the distinctive features that make up a pedestrian such as the edges and gradients. All of this would allow for real-time processing detection which is used for each frame in the prototype. After initialization, the designated video file is opened with a OpenCVs function and frames from the video file is extracted and resized to a width of 400 pixels for better performance and a lack of drop in frames while viewing the video in the results. After the implemented video is processed the detection process starts via HOG with each frame of the video using a function in the HOG library “detectMultiScale()”. This does multiple scans on each resized frame to identify potential pedestrian locations based on learned patterns. Detected pedestrian are then visually indicated by a drawn rectangle around the general area were with the edges of the pedestrian is seen, marking where the OpenCV function “Rectangle()” has spotted and marked a pedestrian. Lastly, the prototype will display the processed video to the user, showcasing the results of the pedestrian detection algorithm. A set of experiments are conducted on the program to record results of different quality-based tests to examine, and stress test the detection process. The experiments will consist of multiple videos and will be captured by a camera that will be positioned on the front dashboard, ensuring that the captured footage closely resembles the visual input experienced during actual driving scenarios. The first set of videos will be captured during daylight, simulating normal driving conditions to assess the model’s detection functionalities. The second set of videos will be recorded at night to examine the model’s ability to detect pedestrians under low-light conditions. Within the imutils package is a function “resize()” that is used to resize images while keeping their aspect ratio. This function will be used in the prototype and will be set to 400 on initial usage for the previously mentioned tests. Following these initial experiments, an additional test will be conducted to evaluate the performance of the pedestrian detection model on less compressed videos with higher resolutions. To simulate this, the resize() function will be adjusted to a width of 700. By increasing the width amount, the images will become less compressed, providing higher resolution and potentially more detailed information for the pedestrian detection model to analyse. This change aims to test the model’s accuracy when presented with higher-quality input data.

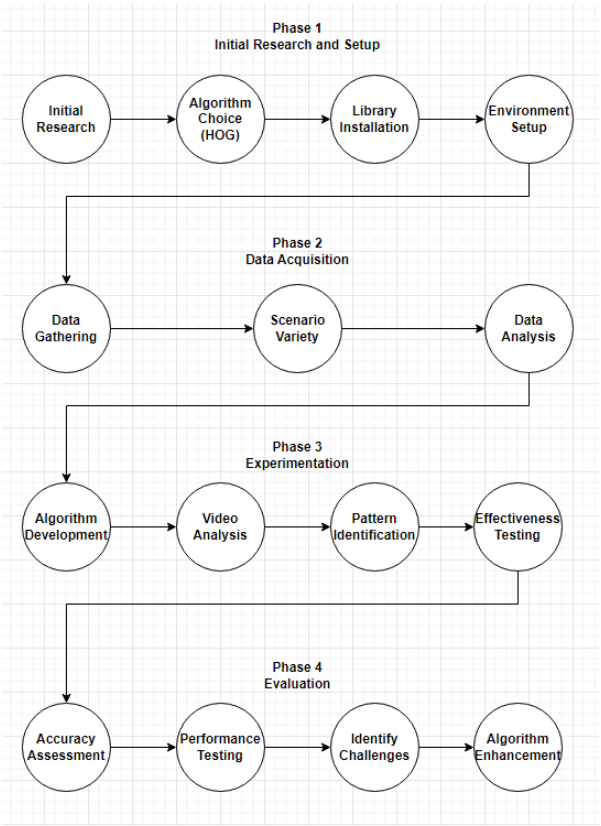

Fig. 1. Research Pipeline

IV. Findings & Discussion of Results

The algorithm implemented in this study has successfully yielded results in pedestrian detection, although with certain

types of limitations and challenges. Throughout the experimental phase, videos were recorded with 1080p quality with a mounted camera on a car dashboard. During this period there were signs that the lighting conditions significantly influenced the detection process. During the day with standard lighting, the algorithm performed well, when it detected a pedestrian, it did not make any mistakes and kept marking them with the desired red square outline. However, despite these successes, false positives were still observed. This suggests that the algorithm excelled in optimal lighting conditions, it still encountered challenges in distinguishing pedestrians from other elements in the environment. This concept was further challenged by taking videos during the evening with low light levels. This resulted in additional challenges as the algorithm struggled to differentiate pedestrians from various sources of light such as streetlights and passing vehicles, leading to a noticeable increase in false positives. However, despite these challenges, the tests revealed that pedestrians were still successfully detected and marked, even in conditions of reduced light level. This reinforces the idea that the algorithm's effectiveness in capturing pedestrians, albeit with an increased likelihood of false positives under low light conditions. Experimenting with the adjustment of the resolution within the project by modifying the "imutils.resize(width)" amount produced varying results. Increasing the width improved the algorithm's ability to detect pedestrians, however, it also led to a higher number of false positives. Conversely, decreasing the resolution resulted in fewer false positives, but the algorithm struggled to detect and mark pedestrians unless they were closer to the camera. This trade-off shows off the importance of balancing resolution settings to optimize detection performance while minimizing errors. After several attempts with different widths, it was found that a range of 300–400 pixels yielded the best results. This range manages to balance detection performance, allowing pedestrians at mid-range distances from the camera to be accurately detected while minimizing the occurrence of false positives. Following these challenges, the project was adjusted to incorporate the consideration for both a minimum and maximum height and width. This adaptation ensures more robust performance across a range of input video qualities and environments. However, moving forward the algorithm is tested using a fixed resolution of 400 pixels for both width and height. Following these adjustments, a series of video quality tests was taken. This involved duplicating a previously tested video, with each duplicate being of progressively lower quality. Specifically, the duplicates were generated in 720p, 360p, and 144p resolutions. These variations in video quality allowed for a close examination of the algorithm's performance across a number of video resolutions, attempting to simulate real-world scenarios where video quality may vary. When these videos were processed through the algorithm to detect pedestrians, it became quickly evident that the video quality played a significant role. It was shown that lower-quality videos presented greater challenges for pedestrian detection. Specifically, a notable tendency appeared where lower-quality videos showed higher instances of false positives and reduced accuracy in pedestrian detection. This displays the sizeable impact of video quality on the algorithm's performance, particularly in the context of Hog's pedestrian detection.

Strength	Weakness
Accurate detection in optimal lightning	Low Light Level Detection
Adaptable Resolutions	Multiple False positives
Effective across video variations	Difficulty detecting pedestrians from light sources

Fig. 2. Strength and Weakness Chart

Comparing the results with those from Pang et al. (2010), where their research utilized HOG and SVM, reveals significant differences. Pang et al. achieved a substantial increase in detection efficiency by reusing features and employing sub-cell-based interpolation on static images [5]. In contrast, my study showed that while the algorithm performed well under optimal lighting conditions, it struggled with false positives in low-light scenarios and varied resolutions.

	Positive	Negative
Positive	True Positive (28)	False Positive (62)
Negative	False Negative (4)	True Negative (12)

Fig. 3. Confusion Matrix Based on a 9-minute video

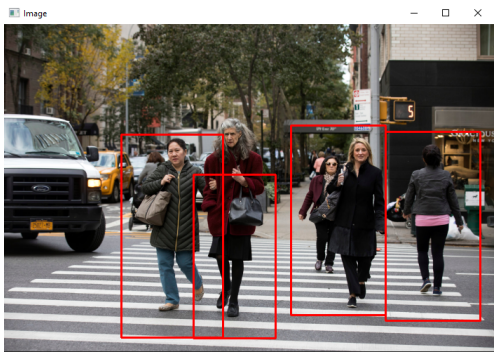

Fig. 4. Tested Image on Detection Process

V. Conclusion

In this study, a pedestrian detection system was developed using Python, OpenCV, and the HOG algorithm. The algorithm is able to process frames to detect pedestrians and mark them in real-time. Experiments and evaluations were conducted under different conditions, such as varying lighting conditions, video quality, and resolutions, all in an attempt

to fully assess the system's performance. These tests aimed to determine how well the pedestrian detection algorithm performed across diverse real life possible scenarios, all to try and identify the strengths and weaknesses along with areas that need improvement. The algorithm exhibited several limitations during the experiments. It struggled with large number of false positives in low-light conditions, where it often misidentified pedestrians due to poor visibility and interference from other light sources such as streetlights and vehicle headlights. Another limitation observed was the system's tendency for false positives even under optimal lighting conditions. While it successfully detected pedestrians, it occasionally marked non-pedestrian objects as false positives. These 2 limitations show that a need for further refinement in distinguishing pedestrians from other elements in the environment to minimize false detections and improve the algorithm's reliability. Additionally, the algorithm's performance varied depending on the complexity of the environment and the presence of occlusions or distractions. In crowded or cluttered scenes, the system struggled to accurately identify pedestrians, leading to potential safety risks if relied upon in such scenarios. In future research within the HOG algorithm a new technique could be formed to further improve the detection system and decrease the rate of false positives possible with the inclusion of SVM similar to the work of Pang et al. [5]. Another area of improvement is in the issue of false positives, particularly in relation to lighting conditions. To reduce this problem, future research could focus on developing algorithms that are more complex to variations of lighting, particularly in low-light environments where visibility is limited. As future research continues to explore the different ways to enhance pedestrian detection systems with the goal of achieving reliable pedestrian detection in diverse lighting conditions to enhance safety on roads, interdisciplinary collaboration will play a crucial role in developing the perfect safety system to avoid future accidents

REFERENCES

[1] S. Zhang, R. Benenson, and B. Schiele, "Citypersons: A diverse dataset for pedestrian detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2017, pp. 3213–3221.

[2] W. R. W. H. Y. Y. Wei Liu, Shengcai Liao, "High-level semantic feature detection: Anewperspective for pedestrian detection," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5187–5196, 2019.

[3] D. Z. X. Y. D. Y. Q. Z. X. W. Yunbo Zhang, Pengfei Yi, "Csanet: Channel and spatial mixed attention cnn for pedestrian detection," *IEEE Access*, vol. 8, pp. 76 243–76 252, 2020.

[4] S. Liu, D. Huang, and Y. Wang, "Adaptive nms: Refining pedestrian detection in a crowd," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2019, pp. 6459–6468.

[5] Y. Pang, Y. Yuan, X. Li, and J. Pan, "Efficient hog human detection," *Signal Processing*, vol. 91, no. 4, pp. 773–781, 2011.

[6] M. Braun, S. Krebs, F. Flohr, and D. M. Gavrila, "Eurocity persons: A novel benchmark for person detection in traffic scenes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 8, pp. 1844–1861, 2019.