Anomaly analysis of convolutional neural networks based pedestrian trajectory determination.

Zoltán Rózsás*; Prof Dr. István Lakatos**

**Széchenyi István Multidisciplinary Doctoral School of Engineering, Department of Automotive and Railway Engineering (rozsas82@gmail.com)*
****Széchenyi István University, Department of Automotive and Railway Engineering (e-mail: drlakatosi@gmail.com)*

Abstract: The accurate determination of pedestrian trajectories is a pivotal aspect of various applications, including road safety, highly automated driving, and autonomous driving. Convolutional Neural Networks (CNNs) have demonstrated remarkable efficiency in spatio-temporal tracking data, making them well-suited for pedestrian trajectory determination. However, the accuracy of CNNs classifications to different kind of objects and their potential impact on trajectory accuracy remain critical concerns. This paper presents an experimental case study about the observed fault that occurred in real environment dataset in the context of CNN-based pedestrian trajectory determination.

1. INTRODUCTION

Anomalies in pedestrian trajectories can arise from diverse sources, such as sudden accelerations, abrupt changes in direction, or interactions with other pedestrians. Therefore, the accurate detection of these is essential for a proper trajectory prediction. Sensing layers like the machine vision-based CNN approach have specific anomalies, that are not properly detected and managed, which can lead to erroneous trajectory predictions and compromise the reliability of downstream applications. This case study is based on real pedestrian movements and crossing-the-road scenarios in the Zalaegerszeg city center. The importance of this research lies in its potential to examine the reliability of pedestrian trajectory determination systems in real-world scenarios. Identified misclassification rate and misrecognition rate help us to better understand the limitations of the method's capabilities. Existing methods of pedestrian tracking and attribute recognition still have not fully addressed two major challenges [1]. The variability of human appearance is very diverse. Different people may look very similar [2], therefore, because of the increased variety, sometimes it is challenging to distinguish simple static objects like a dustbin (Fig 1.). On the other hand, the target may be occluded by neighboring objects or people, making the visual features or attributes noisy and unreliable for recognition [2], [3].

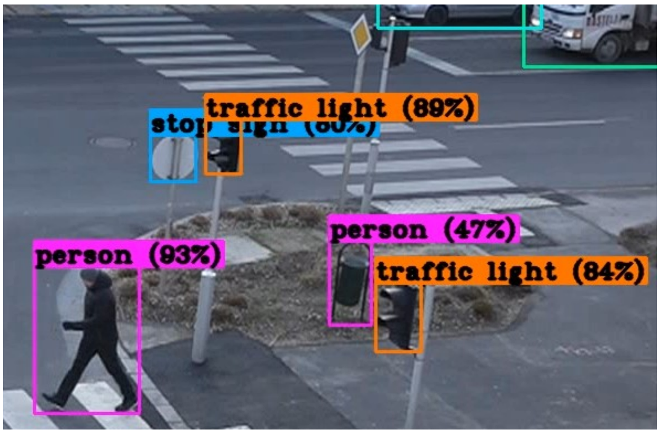

Fig. 1. A static roadside object was classified as a person.

2. BACKGROUND OF THE CASE STUDY

2.1 The location of the experiment

The video images used for the analysis were taken at the intersection of Rákóczi Ferenc út and Arany János út in Zalaegerszeg. (Fig 2.) The viewpoint is located at 7300 mm above the pavement, approximately the height of the traffic lights and the streetlights. The camera location is important for reasons of cost-effectiveness and compactness. With these parameters, the sensors can be placed on existing infrastructure with limitations considering vibrations and movements.

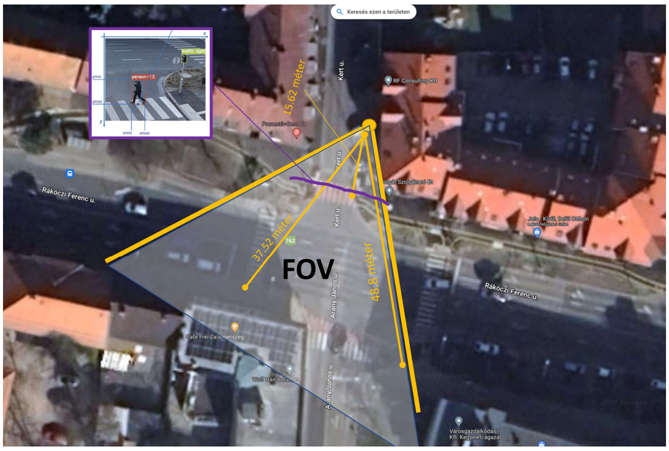

Fig. 2. Location of the experiment – Zalaegerszeg intersection of Rákóczi Ferenc Street and Arany János street

„XVII. IFFK 2023” Budapest
Online: ISBN 978-963-88875-3-5
Paper 05

Editor: Dr. Péter Tamás

2.2 Equipment's

For the recordings used Sony FDR-AX100E (Fig 3.) camera offers a good balance between professional features and user-friendliness. However, mastering its features may require some familiarity with videography techniques and settings adjustments.

Fig. 3. Sony FDR-AX100E used for the base video recordings.

As a 4K camcorder known for its high-resolution video recording capabilities. It features a relatively large Exmor R CMOS sensor for improved low-light performance and image quality, a 20x optical zoom lens for flexible shooting, and manual controls for fine-tuning settings. With a Carl Zeiss Vario-Sonnar T lens, it strikes a balance between professional-grade features and user-friendliness. This camera is widely used in various research studies, for a different kind of trajectory estimation [4] like flame trajectory of a non-vertical turbulent buoyant jet flame.

2.3 Pedestrian trajectory detection by CNN-based models YOLO

The YOLO (You Only Look Once) series of object detection models, developed by Joseph Redmon and later by Alexey Bochkovskiy and the YOLOv4 team, are well-known in the field of computer vision and deep learning for real-time object detection. This is a widely used machine vision tool, that can classify different kinds of objects eg. nine types of skin cancer [5] with a mean average precision score of 88.03%. The model capability comparison is illustrated below (Fig.4.)

The initial version of YOLO, YOLOv1, posed challenges when dealing with images containing multiple small objects of varying sizes. YOLOv2 addressed this issue by incorporating concepts from Faster R-CNN. Subsequently, YOLOv3 emerged as a significantly deeper, faster, and more accurate iteration compared to YOLOv2.

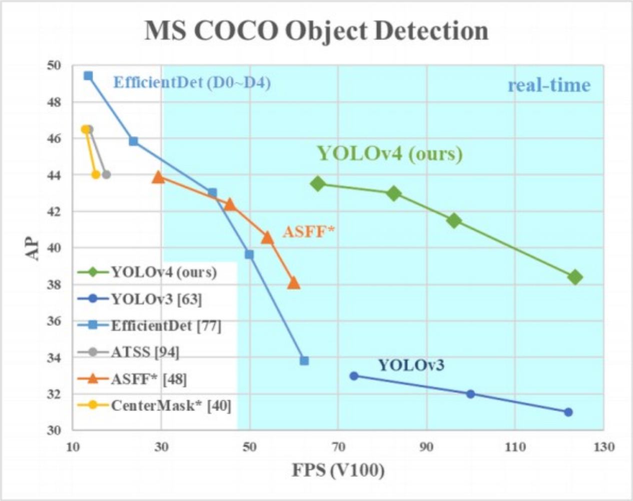

Fig. 4. Comparison of the used YOLOv4 and other state-of-the-art object detectors. YOLOv4 runs twice faster than EfficientDet with comparable performance. Improves YOLOv3's AP and FPS by 10% and 12%, respectively [6].

To enhance real-time object detection accuracy even further, YOLOv4, the fastest version, was introduced in early 2020. Figure 5 illustrates pedestrian detection using YOLOv4.

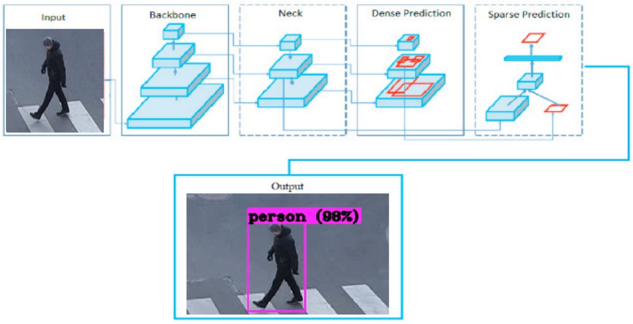

Fig. 5. Pedestrian detection in the city center of Zalaegerszeg by Yolo V4

2.4 Evaluating pedestrian trajectory detection.

CNN-based models, like YOLO, can vary significantly depending on several factors, including the dataset used for training and testing, the complexity of the scenes, and the specific model architecture and implementation. It's important to note that the accuracy of a model can also depend on how well it is fine-tuned for the specific task of pedestrian trajectory detection.

The accuracy of a pedestrian trajectory detection system can be evaluated using various metrics, including:

Detection Accuracy (AP, mAP): The accuracy of detecting pedestrians in a frame. This is typically measured using metrics like Average Precision (AP) or mean Average Precision (mAP).

Trajectory Prediction Accuracy: The accuracy of predicting future pedestrian positions or trajectories. This is evaluated based on metrics like Mean Squared Error (MSE) or root Mean Squared Error (RMSE) between predicted and ground truth trajectories.

Misclassification Rate: This measures the percentage of detected pedestrians that are misclassified as something else (false positives). Lower misclassification rates are desired (Fig. 6.).

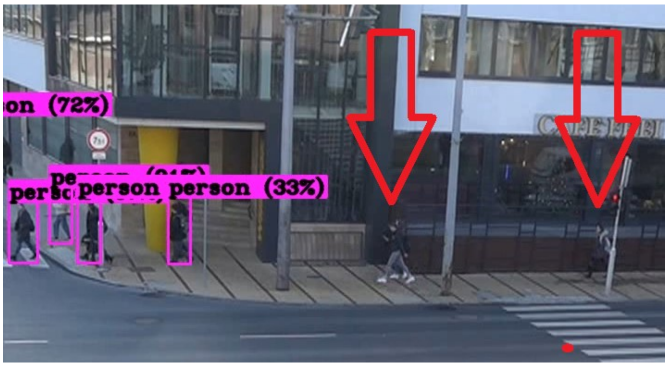

Fig. 6. Miss recognition marked by red arrows Yolo V4.

Recognizing reflections, shadows, or other non-physical entities as real objects can lead to false positives in object detection (Fig 7.). In the context of pedestrian safety, this means that the system may incorrectly detect objects that do not exist in the environment. False positives can trigger unnecessary warnings or actions, which can be disruptive and potentially reduce the system's credibility. E.g. dustbin (Fig. 8) or traffic light (Fig 9.) recognized as a person.

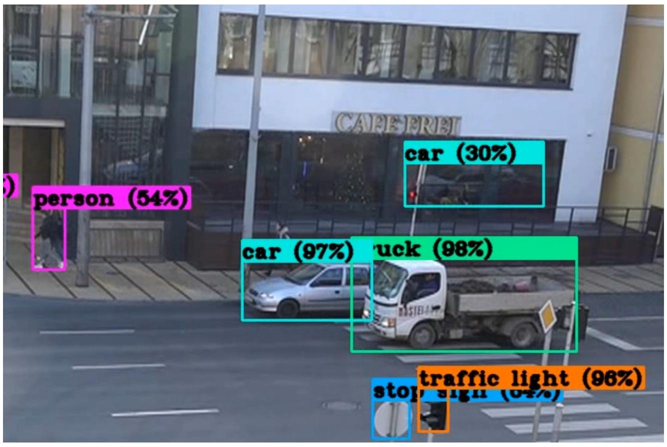

Fig 7. Reflection identified as an object. - Misclassification

Miss Recognition Rate (False Negative Rate): This measures the percentage of actual pedestrians that are not detected (false negatives). Lower miss recognition rates are desired.

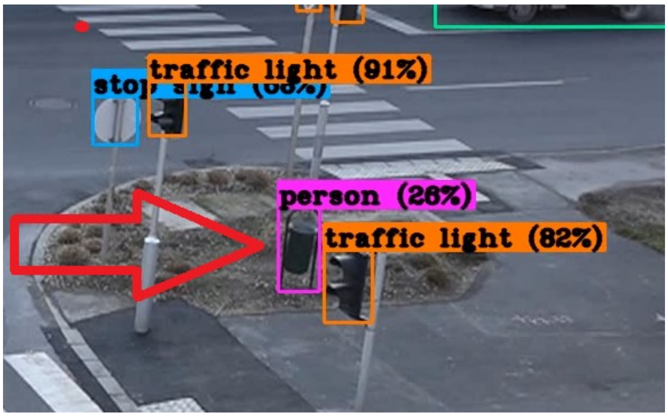

Fig.8. Misclassification marked by red arrows Yolo V4.

The specific values for these metrics can vary depending on the task, the dataset, and the model's capabilities. Achieving high accuracy in pedestrian trajectory detection and prediction is challenging due to the dynamic nature of pedestrian movement, occlusions, and varying scene conditions. Due to the specific conditions and recognition capabilities, I sorted the observed anomalies according to the metrics above.

During the processing of the 10-second video, which contains 252 frames, 1675 object detections were classified against the 2006 Ground truth (GT). Ground truth data provides the correct answers or outcomes for the given dataset, allowing model predictions to be compared to these known values. During the evaluation, typical anomalies were identified, like reflection, misclassification, and miss recognition.

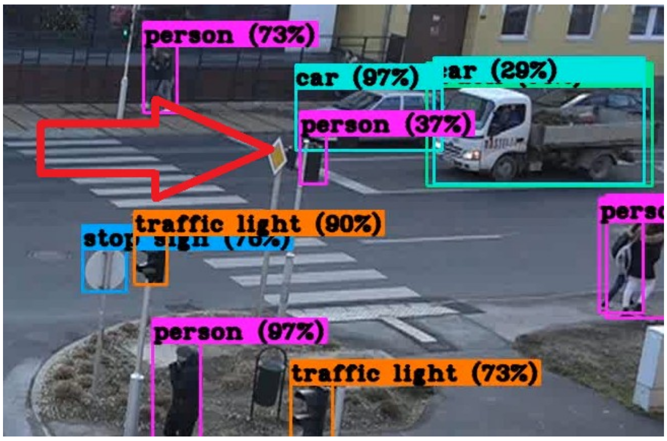

Fig.9. Misclassification marked by red arrows Yolo V4.

3. COMPARISON

To measure the misclassification and miss recognition rate I assign the part of the observed area that is closer to the camera, marked Zone I. (Fig. 10). This area is the basis of the analyzed data. In this “deep short” algorithm that can track the pedestrian, the objects further away are less detectable. The error rate in Zone ll. (AP) is close to the values reported in the literature (44%)[6], however, I use the data sets of Zone l. Although static objects also have an impact on trajectories and therefore on road safety, I consider these errors in the case of misclassification, when static objects are identified as dynamic.

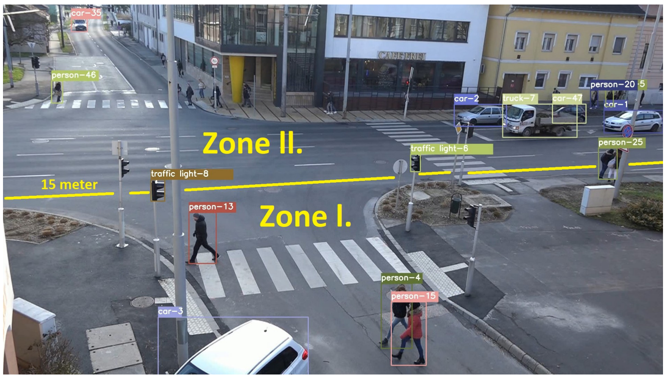

Fig. 10. Assigned Zone I. to the comparison.

Dynamic objects are prioritized in pedestrian safety analysis because they represent a higher safety risk, and accurate detection and tracking of these objects are essential for preventing accidents and improving overall pedestrian safety.

To understanding the behavior of dynamic objects is essential for predicting pedestrian behavior and taking appropriate safety measures. For example, if a pedestrian detection system can accurately identify and track moving vehicles, it can anticipate potential conflicts and take actions such as issuing warnings or triggering automatic braking systems.

3.1 Dataset

Frame by frame numerically compares the objects detected and the Ground truth, which is in machine learning and computer vision, refers to the manually annotated or labeled data. In this dataset, I categorized three groups the detected objects vehicle, pedestrian, and static objects. Each category has two numerical values. One refers to the amount of the recognized object the other is the Ground Truth. Table 1. shows the layout part of the data.

Table 1.

	Nr of Pedestrian GT.	Nr of Recognized Ped.	Nr. Vehicle GT. (P).
Frame 1	2	0	1
Frame 2	2	0	1
Frame 3	3	1	1
Frame 4	3	2	1
Frame 5	3	2	1
Frame 6	3	2	1
Frame 7	3	2	1

3.1 Result of the analysis of the comparison

During the experimental test, the result shows that 83.5% of the GT objects have been detected properly (1675/2006). However, it is also relevant to look at the ratio by category. In the selected area the classified vehicle was parked so this category has up to 99.2% which means 0.8% false negative that is over the expected. On the other hand, the lack of recognition of pedestrians (Fig.11) is 10.4%, which is a weak result considering that we were observed at the more easily detectable closer range and the conditions were ideal.

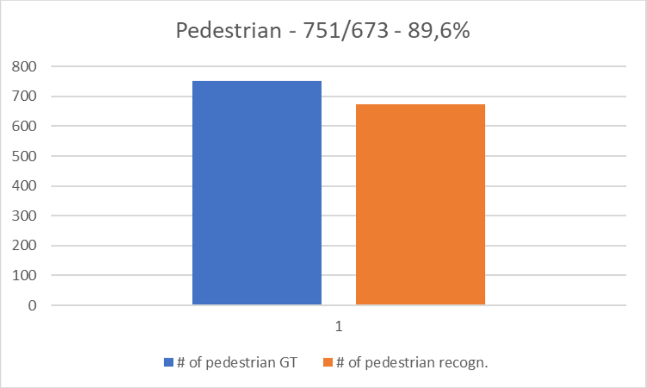

Fig. 11. Detected pedestrian ratio (False negative 10.4%)

The main cause of the misperception was occlusion. In several cases, pedestrians were occluded by static objects such as lampposts, in other cases pedestrians were occluding each other. (Fig. 12.)

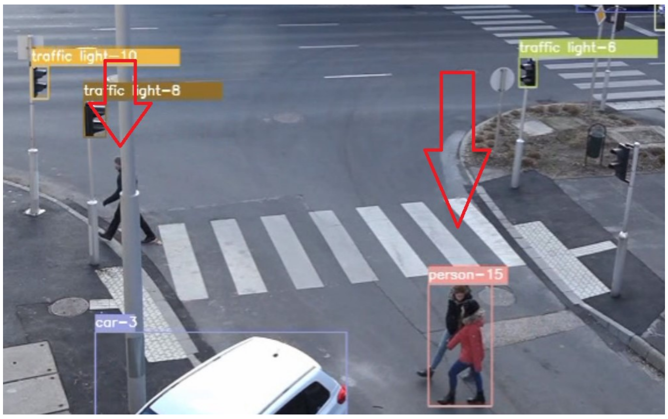

Fig. 12. Occlusion causing false negative misclassification.

6. CONCLUSIONS

In summary, a 10-second video with 252 frames underwent 1675 object detection against the 2006 Ground Truth (GT) dataset. GT data served as a reference for model evaluation, revealing common anomalies like reflection, misclassification, and misrecognition. The study focused on Zone I, the area closest to the camera, vital for pedestrian safety. Dynamic objects, especially pedestrians and vehicles, were prioritized due to their safety impact. The dataset categorized objects into three groups: vehicles, pedestrians, and static objects. Results showed 83.5% accuracy overall, with vehicles achieving 99.2% accuracy but pedestrian detection showed 10.4% false negative.

Occlusion, where static objects or pedestrians obstructed each other, was the main cause of misperceptions. Addressing occlusion challenges is crucial for improving pedestrian detection algorithms and enhancing road safety.

KÖSZÖNETNYILVÁNÍTÁS

A publikációban szereplő kutatást a Széchenyi István Egyetem az Európai Unió támogatásával valósította meg, az Autonóm Rendszerek Nemzeti Laboratórium keretében. (RRF-2.3.1-21-2022-00002)

REFERENCES

[1] Peter Kok-Yiu Wong a, Han Luo a, Mingzhu Wang b, Pak Him Leung a, Jack C.P. Cheng “Recognition of pedestrian trajectories and attributes with computer vision and deep learning techniques” Advanced Engineering Informatics Volume 49, August 2021, 101356

[2] G. Ciaparrone, F. Luque S´anchez, S. Tabik, L. Troiano, R. Tagliaferri, F. Herrera,Deep learning in video multi-object tracking: survey, Neurocomputing. 381 (2020) 61–88, https://doi.org/10.1016/j.neucom.2019.11.023.

[3] X. Wang, S. Zheng, R. Yang, B. Luo, J. Tang, Pedestrian Attribute Recognition: A Survey, ArXiv Prepr. ArXiv1901.07474. (2019). https://arxiv.org/abs/1901.07474.

[4] Wei Gao, Naian Liu, Yan Jiao, Xiaodong Xie, Linhe Zhang a “Flame trajectory of a non-vertical turbulent buoyant jet flame” in Applications in Energy and Combustion Science 7 2021 p12. Volume 7, September 2021, 100039

[5] N Aishwarya, K Manoj Prabhakaran, Frezewd Tsegaye Debebe, M Sai Sree Akshitha Reddy, Posina Pranavee “Skin Cancer diagnosis with Yolo Deep Neural Network” Procedia Computer Science Volume 220, 2023, Pages 651-658

[6] Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao “YOLOv4: Optimal Speed and Accuracy of Object Detection” Computer Science > Computer Vision and Pattern Recognition DOI:2004.10934