Journal of civil engineering and transport **transEngin**
2026, Volume **8** Number **3** **ISSN 2658-1698, e-ISSN 2658-2120**
DOI: 10.24136/tren.2026.011
OPEN ACCESS

PEDESTRIAN CROSSING DETECTION METHOD IN URBAN TRAFFIC USING DEEP LEARNING

Zbigniew **Pamuła** [1]

[1] Silesian University of Technology, Faculty of Transport and Aviation Engineering, Krasińskiego 8, 40-019 Katowice, Poland,
e-mail: zbigniew.pamula@polsl.pl, https://orcid.org/0009-0003-4329-570X

Reviewed positively: 06.03.2026

Information about quoting an article:
Pamuła Z. (2026). Pedestrian crossing detection method in urban traffic using deep learning. *Journal of civil engineering and transport*. 8(3), 159-169, ISSN 2658-1698, e-ISSN 2658-2120, DOI: 10.24136/tren.2026.011

Abstract – Pedestrian crossings, along traffic lights, are an important element of pedestrian safety in urban traffic and suburban roads. Similarly, traffic information is crucial to improving the safety of these vulnerable road users. This paper presents a method for automatic pedestrian crossing detection in car camera images. The method can also be applied using UAV or satellite imagery. A deep learning methodology using a convolutional neural network (YOLOv4) was used to process real-world images. Images obtained from video frames of pedestrian crossings observed at different distances were used to train and test the deep learning model. For test images, which were not part of the training sequence, an accuracy of 77% was achieved for a detection threshold greater than 0.5 and an accuracy of 100% for a threshold greater than 0.01. The developed model can be used to detect pedestrian crossings in various vehicle control applications and in autonomous vehicles. The located objects in the images can also be analyzed for their quality. Crosswalks may be partially dirty or partially worn and may require renovation.

Key words – pedestrian crossing, deep learning, crosswalk detection, camera images.
JEL Classification – C60

INTRODUCTION

Just as road information is crucial for vehicle safety, information about pedestrian amenities (e.g., the presence of sidewalks and crosswalks) is crucial to improving pedestrian safety. To address this need, an automated approach to automatically detecting pedestrian crossings in car camera images was developed. One of the most advanced deep learning methods, the so-called transfer learning using a convolutional neural network (CNN), was used to process real-world images. During the prediction process, the model analyzed the road surface image in the camera image to predict the presence of a pedestrian crossing.

Recent advances in deep learning have led to the development of lightweight CNN models that offer an efficient and effective approach for real-time processing. Deep learning approaches can outperform traditional machine learning methods that rely on hand-crafted features by leveraging transfer learning. In transfer learning, a pre-trained model is used as a feature extractor, and additional layers are added on top to adapt the model to specific tasks. Furthermore, fine-tuning is one of the most common techniques used in transfer learning to adapt the pre-trained model to a new task or dataset.

The specific use case discussed in this paper involves the recognition of pedestrian crossings located within the field of view of a vehicle camera at a distance of no less than a few meters. A schematic diagram of the conducted study is presented in Figure 1.

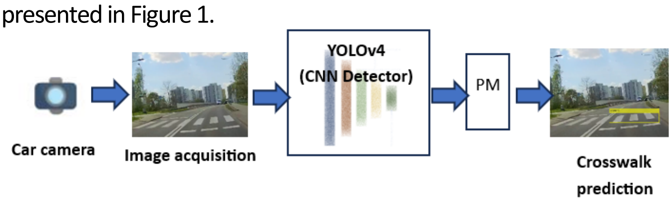


Fig. 1. Schematic diagram of the proposed research model

The next steps, shown in Figure 1, involve acquiring images from the camera video and determining the required resolution. Next, the labelled images are fed to the previously trained YOLOv4 network, and the processing module (PM) processes the network's output, selecting the most accurately recognised object. The model output is an image with a marked pedestrian crossing. A more detailed model development flow chart is presented in Figure 4.

The rest of this paper is structured as follows. Chapter 2 provides a brief overview of publications related to research, including methods for detecting objects in images. Chapter 3 describes the database used in the research and the YOLOv4 neural network used in the proposed pedestrian crossing detection model. Chapter 4 presents the research results, along with their analysis and discussion. Finally, Section 5 provides final conclusions and a plan for future research.

1. RELATED WORKS

Automatic detection of pedestrian crossings in urban environments is an important issue in computer vision and intelligent transportation systems, particularly in the context of driver assistance systems and autonomous vehicles. Early research focused on classical image processing methods that use edge detection, colour analysis, and the Hough transform to identify characteristic road markings. Examples of such approaches are presented in [1] and [2], but these methods had limited robustness to varying lighting conditions, road marking wear, and partial occlusions.

The development of classical machine learning has enabled the use of manually designed feature descriptors, such as HOG [3], in combination with SVM or AdaBoost classifiers. These approaches improved detection performance, but still required careful feature selection and had limited generalisation ability to complex urban scenes.

Significant progress has been made with the use of deep convolutional neural networks. Detection architectures such as Faster R-CNN [4] and YOLOv3 [5] have enabled simultaneous object localisation and classification, achieving high performance in realistic urban traffic conditions. In parallel, semantic segmentation methods based on U-Net [6], DeepLab [7], and NDNet [8] architectures have been developed, which allow accurate separation of pedestrian crossing areas at the pixel level and are particularly useful in systems that require precise road scene analysis.

Public datasets of urban scenes, such as Cityscapes [9] and KITTI [10], played an important role in the development and evaluation of the methods discussed, allowing the training and comparison of deep learning models under realistic conditions. Although these datasets are not exclusively dedicated to pedestrian crossing detection, they are commonly used in conjunction with transfer learning techniques.

Recent research mainly uses deep learning methods to detect horizontal elements of road markings. The authors primarily focus on the accuracy and speed [11] of pedestrian crossing detection, and also link pedestrian crossing recognition to road safety assessments and the condition and quality of road markings. In their article [11], the authors analyzed the speed and accuracy of various pedestrian crossing detection models. They proposed a real-time pedestrian crossing detector called X-CDNet, based on YOLOX. The described detector achieved a very good accuracy of 93.3% and a sufficient detection speed for real-time operation.

In the articles [12-13] the authors presented an automatic system for detecting pedestrian crossings in satellite and aerial images. In these images, crossings are often obscured by trees or buildings, which poses an additional challenge in their recognition. This research could be important for traffic management departments, urban planners, and relevant technology companies to improve road safety.

In [14-16] the authors proposed a defect detection system for road markings based on images of the street scene. These are primarily road markings or pedestrian crossings. The data set used in this study includes five types of defects, including yellow marking lines, white marking lines, directional arrows, pedestrian crossings, and parking guide lines. The authors obtained average precision and AP values slightly above 80%, but the reported results were an average for several recognized road marking elements (classes). Furthermore, only single pedestrian crossing stripes were recognized for pedestrian crossings. Another study [15], related to pedestrian crossings, among others, concerned the detection of road defects, which plays a key role in road maintenance in cities. In [16], a method for automatic detection of road markings, also using deep learning, was presented. The measurement method utilized laser scanning of 3D road surface data. Data on cross slope, cracks, and ruts were analyzed in this way.

The authors of subsequent publications [17-19] linked the problem of recognizing pedestrian crossings with safety in these places.

In [17], an intelligent real-time pedestrian crossing detection system based on video analysis was proposed as a novel safety service for a smart city. The authors' main goal was to improve pedestrian tracking on pedestrian crossings by predicting the future location of the object in real time. The authors of [18] used a deep learning model to identify and regress images of pedestrian crossings based on specific metrics assessing pedestrian exposure to accident events, as well as raw accident data. They used imagery captured using Google Earth.

A system for automatic detection of street features such as traffic lights, pedestrian crossings, and roundabouts was presented in [19]. Deep learning methods were used to detect speed and acceleration patterns for each outlier. Based on this, features were extracted and then classified as locations with traffic lights, pedestrian crossings, city roundabouts, or other features.

In summary, analysis of the literature indicates that deep learning-based methods outperform classical approaches in terms of efficiency and robustness to environmental variability. At the same time, current challenges include reducing the computational complexity of models [17] and improving their performance in real-time, which motivates further research.

Few studies in the literature describe the use of deep learning for pedestrian crossing detection. This is mainly related to the problem of how to detect a crossing, i.e., whether it should be the entire pedestrian crossing [12] or, for example, only individual stripes within the crossing should be detected [14].

The method, proposed in this paper, involves detecting a pedestrian crossing consisting of three stripes. Jet is a new, original idea for detecting a pedestrian crossing directly in front of a moving vehicle. This approach generates more accurate results than detecting individual stripes, which involves the possibility of detecting many other similar objects in the image. Detecting entire pedestrian crossings is important in the case of satellite images and images from unmanned aerial vehicles (UAV), but in the case of a moving vehicle, it is sufficient to detect only pedestrian crossings located a few or a dozen meters in front of the vehicle.

2. MATERIALS AND METHODS

This section presents information about the image database used and a detailed description of the detection model and parameters of the proposed YOLOv4 neural network.

2.1. DATASET

A database of images containing pedestrian crossings was created, extracted from videos recorded with a car camera. The camera was suspended at a height of approximately 1.5 m, and images of the pedestrian crossing were selected so that the distance was not less than 5 m. Images with a resolution of 640x480 pixels were cropped from the video frames. This image resolution was chosen because of the network's training time. Lower image resolutions result in shorter training times.

A total of 100 images were extracted that best depicted pedestrian crossings in the right lane, primarily on two-lane, two-way roads. This set was randomly divided into three subsets. The training set comprised 60% of all images, the test set 30%, and the validation set 10%. This division of the training set ensured correct training of the YOLOv4 network.

Below (Fig. 2), several original images from the training sequence are presented, with the region of interest (ROI) marked on the right. Marking the detection area is essential for training the network. Most pedestrian crossings in Poland were observed to have at least three stripes. Therefore, such areas were selected for training the deep learning neural network.

To train the network, a detection area (annotated image) must be marked on the original images. In our case, the detection area is a section of a pedestrian crossing on the right lane, located a few to several meters ahead of the moving vehicle and containing three stripes. To increase the amount of training data and improve its accuracy, data augmentation was performed, consisting of the following image processing operations:

- Randomly change the hue, saturation and brightness of the image in the HSV colour space ((Hue, Saturation, Value)),
- Randomly flip the image horizontally,
- Randomly scale it by 10%.

Data augmentation was not applied to the test and validation data. The test and validation data are representative of the original data and were left unmodified for objective evaluation.

Each of the training images was subjected to transformations, the effects of which are presented in Figure 3.

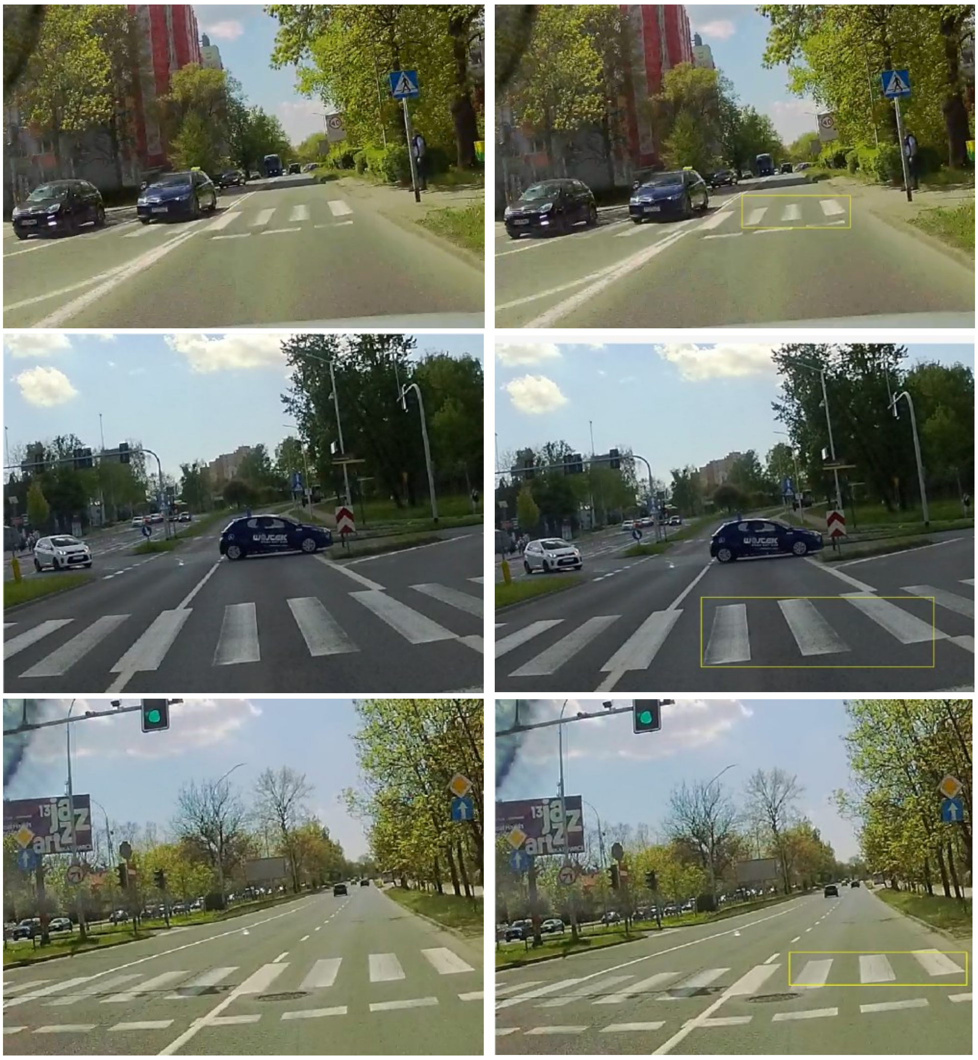


Fig. 2. Sample images from the training sequence with selected objects (**right**)

2.2. Network overview

The YOLOv4 network (You Only Look Once ver4) was used to detect crosswalks. This deep learning network, due to its speed, can be used for real-time object detection in images for image recognition systems. YOLO is a series of algorithms in computer vision that use neural networks to predict the location of rectangular boundaries and classification probabilities in a single image pass. YOLO has found widespread use in real-time object detection in images and videos.

The primary author of the original YOLO was Joseph Redmon [20]. The design of the first version of the object detector was published in 2016 as part of a custom framework called Darknet. Subsequent versions, such as YOLOv2 and YOLOv3, were more accurate and faster, but still struggled to detect small or densely packed objects, as well as those significantly occluded. YOLOv4 [21] was chosen for this study, as it is fast enough to solve our problem. The YOLOv4 network consists of 74 layers, including the input layer, contains 21 convolutional layers that, after training, transform the image using filters.

Fig. 3. Example of the image after augmentation

A pre-trained Tiny YOLOv4 network was used to detect pedestrian crossings on the COCO (Common Objects in Context) dataset. There is a so-called lightweight version of the YOLOv4 network with fewer network layers. The COCO database is a large-scale text detection, segmentation, and captioning dataset created by Microsoft. Using a network pre-trained for object recognition, the network can be trained to recognize other objects in a short time (e.g., within an hour or a few hours) on a standard personal computer.

To obtain a reliable evaluation of the proposed model, the network was trained repeatedly with varying parameters. Ultimately, the parameter values that yielded the best results were determined. The hyperparameter values used in the proposed detection model are presented in Table 1.

Table 1. Model hyperparameter values

Parameter	Value
Learning rate	0,001
Input size	416x416
Batch size	8
Max epoch	50
Optimizer	adam

Maintaining compliance with the basic network parameters requires processing training images with a resolution of 640x480 using appropriate functions (transform, preprocessData) to a size of 416x416 pixels and estimating the coordinates of detection objects (bounding boxes).

The "adam" (adaptive moment estimation) solver was used to train the neural network. “Adam” is a stochastic solver that is most often used to train YOLOv4 networks.

In this study MATLAB R2025a software (Academic license) and a personal computer with an Intel Core i5 processor and 32GB of RAM were used.

2.3. CROSSWALK DETECTION MODEL

Figure 4 shows the steps needed to create the model.

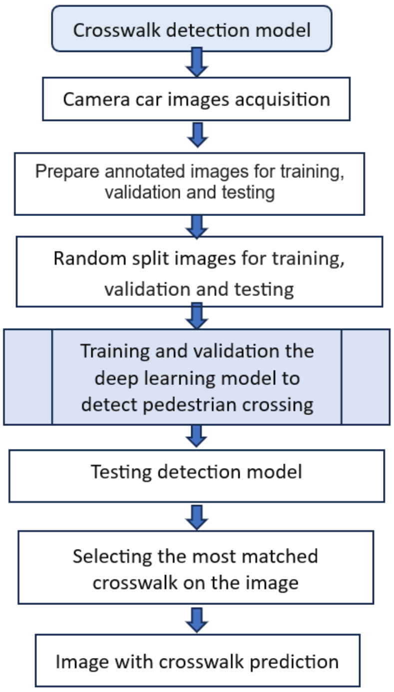


Fig. 4. Flow chart of steps in the proposed methodology

This flowchart shows the sequential steps. First, images containing pedestrian crossings were selected from a video recorded with a dashcam. The next step involved marking an ROI (Region of Interest) on all images—in our case pedestrian crossings. After this operation, the set of prepared images was randomly divided into three parts: a training set, a validation set, and a test set. The training dataset was enlarged using an augmentation operation.

The most important element of the entire model is the training and validation process of the YOLOv4 neural network. The quality and accuracy of the detector depend on the accuracy of the training. The network was trained repeatedly for 50 epochs. The training time ranged from 50 to 120 minutes, depending on the values of the network parameters.

3. RESULTS AND DISCUSSION

The trained YOLOv4 network was tested on 30 images that were not part of the training set. In most of the images, the network detected multiple areas that contained pedestrian crossings. To eliminate overlapping areas, the one with the highest Scores value was selected. This value is displayed on the border of the detected area. In all tested images, the pedestrian crossing area was detected correctly. Example detector results are shown in Figure 5.

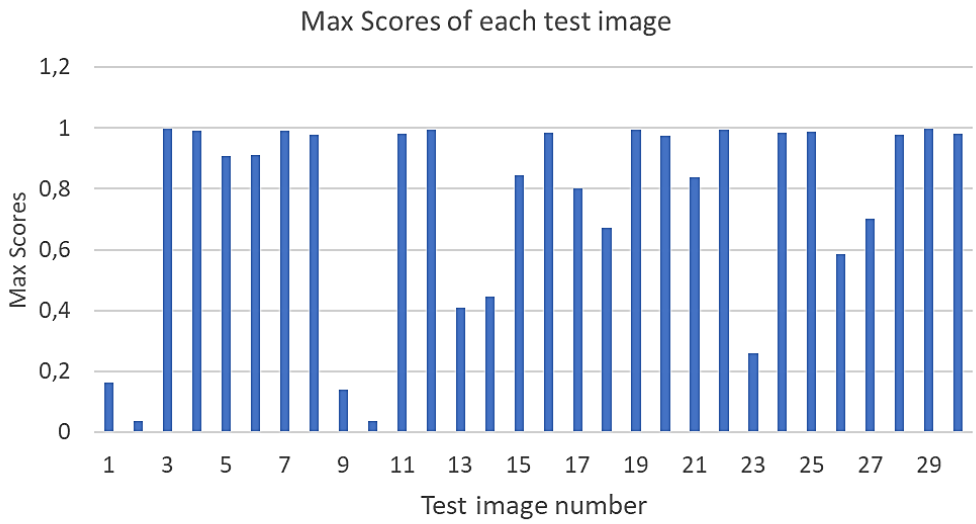
As mentioned above, when detecting a fragment of pedestrian crossing, the detector often identified areas that overlapped. These areas were detected with varying accuracy. For objects detected in their entirety, such as vehicles or people, the value of the Scores detection parameter is crucial. The maximum value of this parameter is 1, when the detected object is completely within the bounding box. If the box is shifted relative to the detected object, the Scores value is less than 1. In our case, this is not as significant as the detected pedestrian crossing fragment should contain at least three stripes, not necessarily entirely within the bounding box. In our case, a value above 0.01 produced an acceptable detection result for this study (Fig. 6).

Fig. 5. Crosswalk detector results

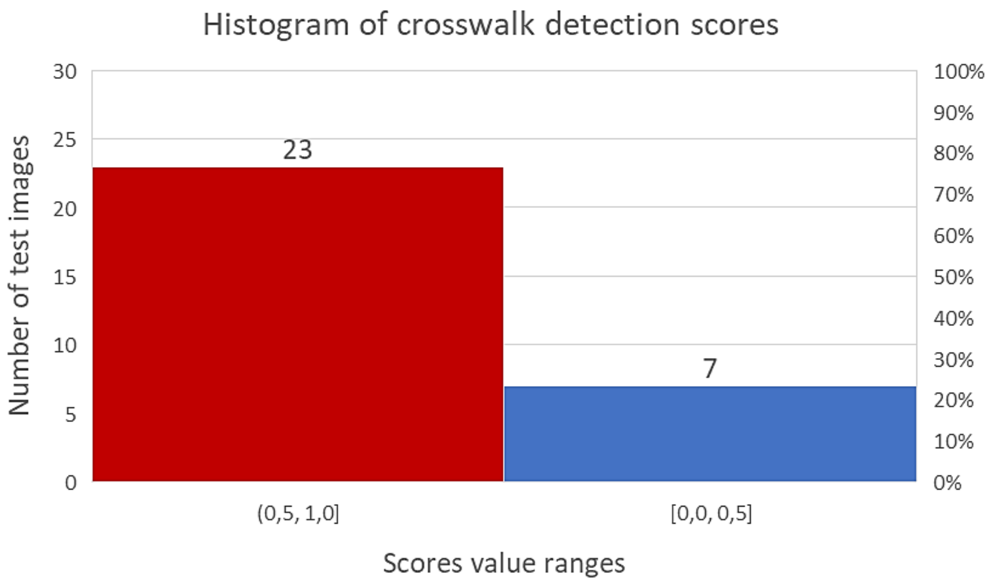
The graphs in Figure 6 show the Scores results for each of the 30 test images.

Graph (1) in Figure 6 shows the best crosswalks detection results for each test image. Most values are above 0.9, which corresponds to near-perfect detection. The histogram on the right side of Figure 6 shows that of the 30 test images, 77% (23 images) were recognized with Scores above 0.5, while the remaining 23% (7 images) were recognized with Scores below 0.5. These results can be considered very good in our case, because, as previously mentioned, results below 0.5 are also sufficient for our study.

The pedestrian crossing detection results for the test images numbered 1 to 6 with the scores of the graph on the left side of Figure 6 can be seen in the images in Figure 5. Image number 1 (Fig. 5) corresponds to number 1 in the graph, image number 2 in the graph corresponds to number 2, and so on.



(**1**)



(**2**)

Fig. 6. Maximum crosswalk detection results for each test image (**1**), **histogram** (**2**)

Figure 7 shows sample test images in which more than one area was detected. Due to the way the pedestrian crossing section was selected (on the right side of the road, three stripes), the proposed detector often detected more than one area.

In most of the test images, two or more objects containing pedestrian crossings were detected. Detection areas were detected correctly, although they often shifted from the reference areas established prior to testing. Selecting a single detected area meant keeping the one with the highest Scores value.

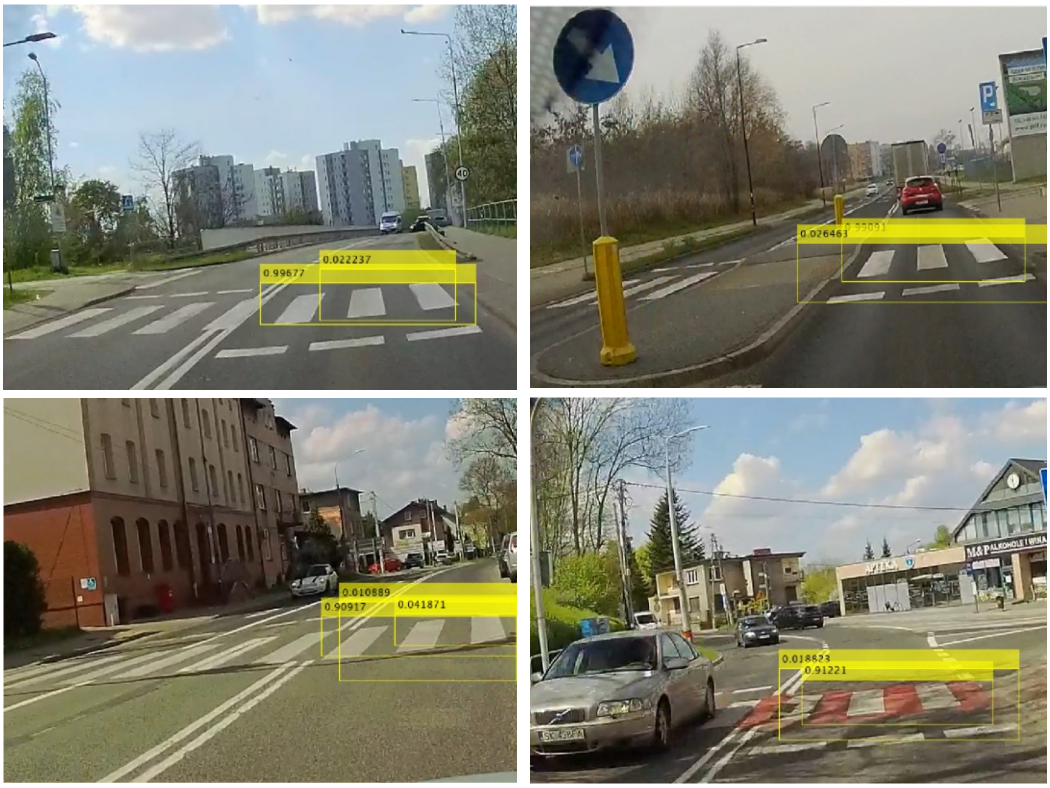


Fig. 7. Sample test images showing the detection results

CONCLUSIONS

This article presents the model of pedestrian crossing detection in images obtained from a car camera. The fast YOLOv4 neural network was used for object recognition. This network is characterized by good object detection results and high speed, which allows its use in real-time detectors.

Pedestrian crossings were recognized in all tested images, the result being accurate in 77% of them, while in the remaining 23%, the detection field was slightly shifted from the expected value. However, pedestrian crossing stripes were recognized in all images. The method's primary advantages are its speed and repeatability. The stripes were detected in all the test images. The disadvantage of the method is the varying accuracy of detection – the stripes were not always detected in their entirety, i.e. three at a time, as intended.

In this study, it was assumed that recognition of at least three stripes in a pedestrian crossing would be sufficient. Future research will examine other variants of objects, such as individual pedestrian stripes or entire pedestrian crossings on both sides of the road. A larger number of images will also be used to train the network more accurately. Currently, networks like YOLO can also detect objects with different shapes. This approach would enable accurate marking of each pedestrian lane.

Research on pedestrian crossing detection will continue. The results can be used in vehicles (not just autonomous vehicles) and also to assess the quality of pedestrian crossings, for example, in terms of dirt or wear (abrasion).

Research will be conducted on the quality of road markings and the results will be presented in future publications.

METODA WYKRYWANIA PRZEJŚĆ DLA PIESZYCH W RUCHU MIEJSKIM Z WYKORZYSTANIEM GŁĘBOKIEGO UCZENIA

Przejścia dla pieszych, obok sygnalizacji świetlnej, stanowią ważny element bezpieczeństwa pieszych w ruchu miejskim, a także na drogach podmiejskich. Podobnie jak informacje drogowe mają kluczowe znaczenie dla poprawy bezpieczeństwa tych wrażliwych użytkowników dróg. W artykule przedstawiono opracowaną metodę automatycznego wykrywania przejść dla pieszych na obrazach z kamery samochodowej. Metoda może być również zastosowana z użyciem zdjęć uzyskanych UAV lub satelitarnych. Do przetwarzania obrazów rzeczywistych wykorzystano metodykę głębokiego uczenia z wykorzystaniem splotowej sieci neuronowej (YOLOv4). Do trenowania i testowania modelu głębokiego uczenia wykorzystano zdjęcia wycięte z klatek filmu z widokiem pasów dla pieszych obserwowanych z różnych odległości. Dla obrazów testowych, nie będących elementami ciągu uczącego, uzyskano dokładność 77% dla progu nakładania większego od 0,5 i dokładność 100% dla progu powyżej 0,01.

Opracowany model może być stosowany do wykrywania przejść dla pieszych w różnych zastosowaniach sterowania pojazdami oraz w pojazdach autonomicznych. Zlokalizowane obiekty na obrazach mogą być również poddane analizie pod kątem ich jakości. Pasy mogą być częściowo zabrudzone lub częściowo zużyte i wymagające renowacji.

Słowa kluczowe: głębokie uczenie, przejście dla pieszych, obrazy z kamery, wykrywanie przejść dla pieszych.

AUTHOR CONTRIBUTIONS

– conceptualization, Z.P. (Zbigniew Pamuła);
– methodology, Z.P.;
– software, Z.P.;
– validation, Z.P.;
– formal analysis, Z.P.;
– investigation, Z.P.;
– writing - original draft preparation, Z.P.;
– writing -review and editing, Z.P.;
– visualization, Z.P.;
– supervision, Z.P.;
– project administration, Z.P.

All authors have read and agreed to the published version of the manuscript

Conflicts of Interest: The author declare no conflicts of interest.

REFERENCES

[1] Zhu C., et al. (2023). Zebra-crossing detection based on cascaded Hough transform principle and vanishing point characteristics, *Open Computer Science*, 13(1). https://doi.org/10.1515/comp-2022-0260.

[2] Wu X. -H., Hu R. Bao Y. -Q. (2019). Block-Based Hough Transform for Recognition of Zebra Crossing in Natural Scene Images, *IEEE Access*, 7, 59895-59902. https://doi: 10.1109/ACCESS.2019.2915639.

[3] Sun Z., Caetano E., Pereira S., Moutinho C. (2023). Employing Histogram of Oriented Gradient to enhance concrete crack detection performance with classification algorithm and Bayesian optimization, *Engineering Failure Analysis*, 150, ISSN 1350-6307. https://doi:10.1016/j.engfailanal.2023.107351.

[4] Ren S., et al. (2017). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6), 1137–1149. https://doi: 10.1109/TPAMI.2016.2577031.

[5] Deng L., Li H., Liu H. et al. (2022). A lightweight YOLOv3 algorithm used for safety helmet detection, *Sci Rep* (12), 10981. https://doi: 10.1038/s41598-022-15272-w.

[6] Ronneberger O., Fischer P., Brox T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention (MICCAI), *Lecture Notes in Computer Science*, 234–241. https://doi: 10.1007/978-3-319-24574-4_28.

[7] Chen L.-C., et al. (2018). DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848. https://doi: 10.1109/TPAMI.2017.2699184.

[8] Yang Z., et al. (2021). NDNet: Narrow While Deep Network for Real-Time Semantic Segmentation, *IEEE Transactions on Intelligent Transportation Systems*, 22(9), 5508-5519. https://doi: 10.1109/TITS.2020.2987816.

[9] Janosovits J. (2022). Cityscapes TL++: Semantic Traffic Light Annotations for the Cityscapes Dataset, 2022 Int. Conference on Robotics and Automation (ICRA), Philadelphia, PA, USA, 2569-2575. https://doi: 10.1109/ICRA46639.2022.9812144.

[10] Chevrin G., et al. (2022). Visual-Inertial Fusion on KITTI using MSF-EKF, 2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV), Singapore, 35-40. https://doi: 10.1109/ICARCV57592.2022.10004344.

[11] Lu X., Xue Y., Wang Z., Xu H., Wen X. (2024). X-CDNet: A real-time crosswalk detector based on YOLOX. *Journal of Visual Communication and Image Representation*, 102, ISSN 1047-3203. https://doi.org/10.1016/j.jvcir.2024.104206.

[12] Luttrell J., Zhang Y., Zhang Ch. (2024). Automatically detect crosswalks from satellite view images: A deep learning approach with ground truth verification, *International Journal of Transportation Science and Technology*, 16, 165-176. https://doi.org/10.1016/j.ijtst.2024.01.006.

[13] Russon D., et al. (2025). Evaluating pedestrian crossing safety: Implementing and evaluating a convolutional neural network model trained on paired aerial and subjective perspective images. *Heliyon*, 11(4).

[14] Chen T., et al. (2024). Road marking defect detection based on CFG_SI_YOLO network, Digital Signal Processing, (153). https://doi.org/10.1016/j.dsp.2024.104614.

[15] Jia Y., Yang Ch., Gao Y. (2026). LL-YOLO: A lightweight pavement disease detection algorithm in low-light environments, *Measurement*, 258(D). https://doi.org/10.1016/j.measurement.2025.119494.

[16] Zhang D., Xu X., Lin H., Gui R., Cao M., He L. (2019). Automatic road-marking detection and measurement from laser-scanning 3D profile data, *Automation in Construction*, 108. https://doi.org/10.1016/j.autcon.2019.102957.

[17] Yang Y., Lee S., Lee J. (2022). Video analytics-based real-time intelligent crossing detection system (RICDS): Killer app for edge computing, *Future Generation Computer Systems*, 133, 84-94. https://doi.org/10.1016/j.future.2022.03.013.

[18] Mussone L., Hassan O. (2025). An analysis of pedestrian crossings through deep learning models and crash data. *Transportation Research Interdisciplinary Perspectives*, 31, ISSN 2590-1982. https://doi.org/10.1016/j.trip.2025.101449.

[19] Munoz-Organero M., et al. (2018). Automatic detection of traffic lights, street crossings and urban roundabouts combining outlier detection and deep learning classification techniques based on GPS traces while driving, *Computers, Environment and Urban Systems*, (68), 1-8. https://doi.org/10.1016/j.compenvurbsys.2017.09.005.

[20] Redmon J., et al. (2016). YOLO: real-time object detection, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 779-788. https://doi.org/10.1109/cvpr.2016.91.

[21] Bochkovskiy A, Chien-Yao Wang, and Hong-Yuan Mark Liao. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection, *arXiv preprint arXiv*: 2004.10934. https://doi.org/10.48550/arXiv.2004.10934.