Bitlis Eren Üniversitesi Fen Bilimleri Dergisi
BİTLİS EREN UNIVERSITY JOURNAL OF SCIENCE
ISSN: 2147-3129/e-ISSN: 2147-3188
VOLUME: 13 NO: 4 PAGE: 885-895 YEAR: 2024
DOI: 10.17798/bitlisfen.1432965

A New Approach to Automatic Detection of Tactile Coating Surfaces with Deep Learning

Abdil KARAKAN[1*]

[1]*Afyon Kocatepe University, Dazkırı Vocational School, Department of Electrical, Afyonkarahisar, Türkiye*
(ORCID: 0000-0003-1651-7568)

Keywords: Tactile coating, Deep learning, Blind, Real-time detection, YOLO.

Abstract
In this study, tactile coating surfaces of visually impaired individuals were detected using the deep learning method. For this detection, 4 of the You Only Look Once (YOLO) architectures, one of the best deep learning methods, were used. No ready data set was used in the study. A unique and new data set was prepared for the study. For the data set, 6278 images were taken from tactile coating surfaces. Images for real-time applications were obtained from many different environments. The tactile coating surfaces in the pictures were labelled separately. A total of 9184 tags were made. The dataset was implemented in YOLOv5, YOLOv6, YOLOv7, and YOLOv8 architectures. The highest accuracy was achieved in the YOLOv8 architecture with an accuracy rate of 97%, F1-Score of 0.940, and mAP@.5 of 0.977. The model was applied with k-fold cross-validation to evaluate performance measurements. In order for the study to be used in real-time, the frame per second (FPS) was increased to 150.

1. Introduction

According to World Health Organization data, 1 billion people in the world have vision problems. Of this, 123.7 million are unidentified visual defects, 65.2 million are cataracts, 6.9 million are glaucoma, 4.2 million are corneal opacity, 3 million are diabetic retinopathy, 2 million are chronic conjunctivitis, and 826 million are unidentified presbyopia-related visual disorders [1].

People with different degrees of visual impairment are closely affected by physical environmental conditions as well as economic and socio-cultural conditions. Urban space arrangements that do not meet the needs of visually impaired people often result in these people being confined to their homes. Visually impaired individuals can play an active role in society by providing limited vision with modern technology, the use of a cane or a guide dog [2]-[5].

Sensible surfaces were first developed by Seiichi Miyake in Japan in 1965. They are called “Tenji blocks” [6]. They were first used in the city of Okayama in 1967 [7]. Then they gradually spread to other cities. The low cost of tactile surfaces has made them considered the most effective system for the visually impaired. Sensible surfaces are basically of two types: stimulating blocks and guiding blocks [8], [9]. The first use of tactile surfaces was by arranging dome-shaped points parallel to each other. Afterward, different tactile surfaces arranged in a zigzag shape, oval shape and bar/line shape were used. Sensible surfaces, usually yellow or red, are preferred for the circulation of the visually impaired [10-12]. They are used on sidewalks, subways, underpasses, bus and train stations, and other public areas. Although most people think that they are used for design purposes, they play a major role in the daily lives of visually impaired individuals [13]-[20].

Many studies have been done to aid the circulation of visually impaired individuals. With the development of technology, systems that can detect obstacles in advance and guide disabled individuals have been designed [20]-[30].

Einloft aimed to create a system that can detect indoor spaces such as shopping malls,

[*]Corresponding author: *akarakan@aku.edu.tr*
Received: 06.02.2024, Accepted: 03.10.2024

subways, and bus terminals in order to help disabled individuals. The study was carried out in two stages. In the first stage; the gray-level co-occurrence matrix method used in image classification was used. In the second stage; Support vector machines, one of the supervised learning methods, were used. Although the initial results were promising, no numerical values regarding study accuracy were given [31].

Ghilardi has detected tactile parquet surfaces in real time to help visually impaired individuals. In this study, edge detection algorithms such as Canny, Lap-lace, and Sobel were used. In this study, they obtained a new approach by combining decision trees. There are 521 photographs in the created data set. Photographs were obtained with a smartphone attached to the person's waist at a height of one meter and at a 45-degree angle. There are 320 photographs of tactile parquet surfaces. There is no tactile parquet surface in the 201 photographs in the data set. In this study, the system operates at 16 FPS. The accuracy rate of the developed system is 88.48% [32].

Tactile parquet surfaces were detected using Jie image processing algorithms. Multicolor image segmentation and Kirsch edge detection algorithm were used in the system. Using the Hough transformation orientation, the tactile parquet surface was obtained as a strip. No numerical value is mentioned regarding the operating accuracy of the system, which has a 2 FPS operating speed and can be used portablely [33].

Asami and Ohnishi created a system that helps visually impaired individuals find themselves at a pedestrian crossing. The system consists of four parts. These consist of a USB camera integrated into the glasses, a vibrating wristband attached to the wrist, a notification system that gives warnings in the visually impaired alphabet integrated into the visually impaired cane, and a computer used to process the data coming from the USB camera. In this study, pedestrian crossings are detected using the normalized cross-correlation method in photographs taken from a USB camera. When a pedestrian crossing is detected, the user's wristband vibrates, allowing the user to stop in front of the pedestrian crossing. The developed system detects pedestrian crossings with 86.7% accuracy [34], [35].

Aktaş and his colleagues detected tactile parquet surfaces in real-time in order to help visually impaired individuals. In their study, they used YOLOv2, YOLOv3, and YOLOv3 DenseNet architectures. They prepared the data set themselves. Photographs were obtained from places that are frequently used in daily life, such as libraries, subways, parks, walking paths, and streets in various districts and regions of Istanbul. The captured photographs were converted to 1024 × 768 in order to be used quickly and conveniently in the system. The larger the data set, the higher the accuracy rate of the study. For this purpose, a data augmentation process was applied. The data set initially consists of 1200 photographs. Then, the data set was replicated with the OpenCV Library. As a result of the data duplication process, 4580 data points were accessed in the system. During the data duplication process, operations such as mirroring, whitening, and rotation were performed. The dataset was used individually in three different YOLO architectures. The results were compared. The best results were achieved with the YOLOv3 DenseNet architecture. It has been understood that YOLOv3 DenseNet architecture is better than other models in detecting tactile parquet surfaces with 89% F1-score, 92% average sensitivity and, 81% IoU values [36], [37].

As a result of the literature review, it is understood that the use of previously made systems in daily life is quite difficult. These difficulties are systems with low FPS values. Real-time detection cannot be made with a low frame rate. In addition, previous studies have carried out the detection of tactile parquet surfaces both indoors and outdoors with only one system. Shopping mall, metro, etc. A separate system is created for interior spaces such as walkways and streets, and a separate system for exterior spaces such as walkways and streets. Therefore, the user must use extra equipment such as a computer, cane, and wristband in his backpack to use the system. In the studies conducted, there are problems such as the inability to complete the studies due to lack of data. Unlike the studies in the literature, the system used in this study performed real-time detection at 150 frames per second. With this system, tactile parquet surfaces were detected both indoors and outdoors without the need for extra equipment such as canes, wristbands, or computers. For this reason, a lot of photographs were used while preparing the data set. The photographs that make up the data set were collected from different places and at different times. Thus, a data set closer to real life was created. With the created data set, a system that works very well in real life has been realized.

In this study, a model that can detect tactile parquet surfaces in photographs, videos, and real-time operating systems that can be used in assistive technologies for individuals with visual impairments was developed. Considering the applicability and suitability of deep learning methods in the field of image processing, deep learning methods and image processing algorithms were mainly used in this study. When it comes to object detection based on deep learning, the YOLO method gives very good results in real-time detection. A data set consisting of 6278 labeled photographs was used as the data set within the scope of this study. In the experiments conducted on the data set, tactile parquet surfaces were detected with an F1 score of 89%, an average sensitivity of 92%, and an IoU value of 81%. The study has three important contributions. In this study carried out, a new and large data set was created in this field, and tactile parquet surfaces were detected by using the YOLO method on this data set without depending on a specific camera angle. A new model with high-performance values has been developed. Deep learning methods have been used in the field of tactile parquet surface detection.

2. Material and Method

In this study, a model that can detect tactile parquet surfaces in photographs, videos, and real-time operating systems that can be used in assistive technologies for individuals with visual impairments was developed. The study was carried out in three stages. In the first stage, the data set was prepared. For this purpose, tactile coating surfaces found in real life were used. Thus, when the study was implemented, a higher rate of compliance was achieved in real life. In the second stage, the data set duplication process was applied. As the data set increased, the accuracy rate of the system increased. The highest accuracy was achieved by using the created data set in four different architectures. Since the work will be real-time, the FPS rate must also be very good. In the last stage it has been implemented in real life. Figure 1 shows the general structure of the study carried out for the detection of tactile coating surfaces.

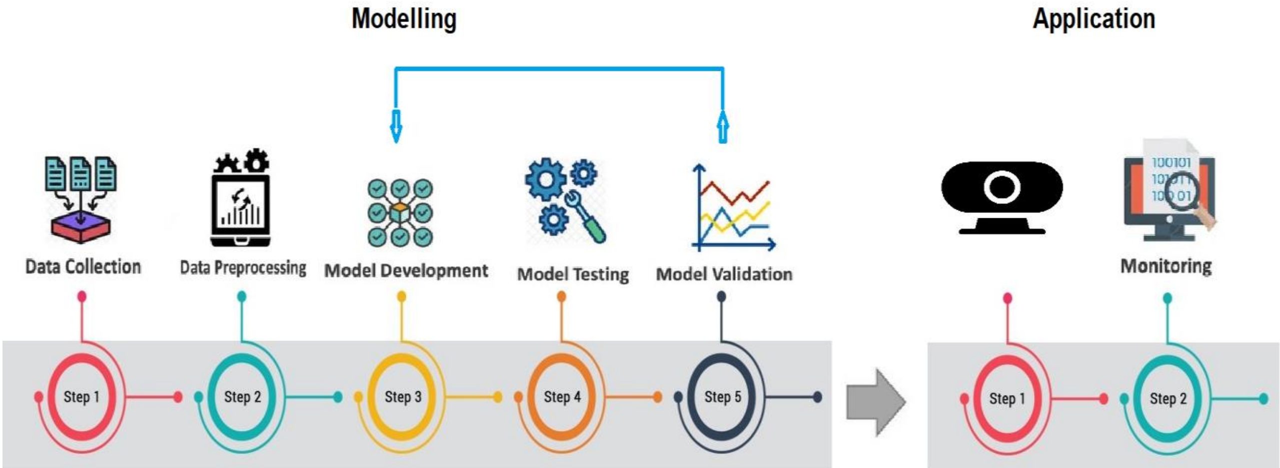
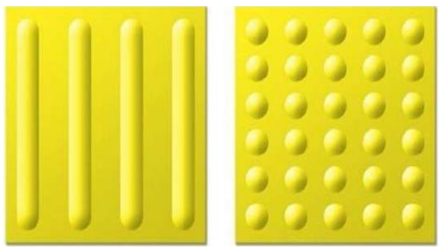


Figure 1. The general structure of the study was carried out for the detection of tactile coating surfaces.

2.1. Data Set

Tactile surface indicators should be arranged to indicate access to facilities along the route and warn of possible dangers. Among these indicators, the one in the form of a bar/line is called the guiding surface, and the one with dots is called the warning surface. Indicators in the form of bars/lines provide guidance in finding direction. Dotted indicators serve as warnings by placing them in potentially dangerous areas such as stairs, pedestrian crossings, and train/subway platforms. In short, bar/line indicators mean “go” and dotted indicators mean “stop” [38], [39].

While preparing the data set, tactile surface samples of different types and from different environments were taken. For this purpose, 6278 photographs were taken. In these photographs, 9814 tactile surfaces were labeled. Figure 3 shows 9 different tactile surface examples from the data set. Since real-time detection will be made in this study, photographs were taken from as real environments as possible. Figure 2 shows examples of tactile surface parquet.



a) **b)**

Figure 2 a) Guiding tactile surface **b)** Stimulating tactile surface.

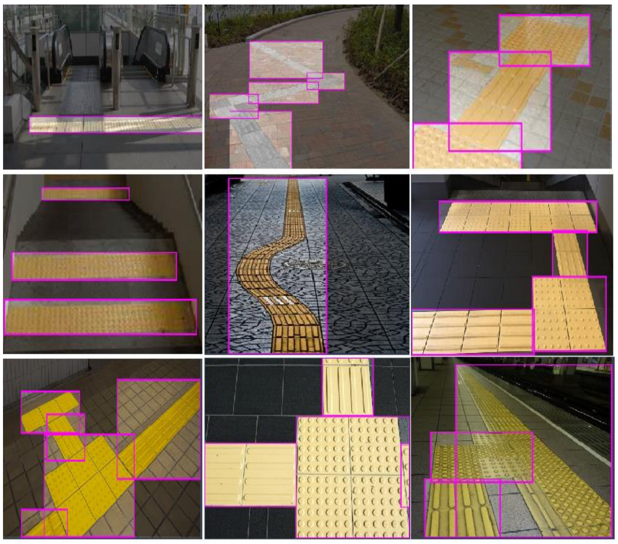


Figure 3. Labeling of different photographs used in the data set.

2.2. YOLO Architecture

Deep Learning is one of the sub-topics of the field of machine learning and is also the most current topic. Deep Learning methods are used as artificial neural network algorithms inspired by the structure of the human brain. Deep Learning is part of a broader family of machine learning methods based on learning data representations as opposed to task-specific algorithms. Deep Learning, along with studies in the field of machine learning, has also pioneered the expansion of the scope of artificial intelligence. Compared to superficial learning, the concept of deep learning has the advantage of building deep layers to reach more abstract information. Deep learning-based methods are used successfully in many fields. One of these is automatic object detection. In this study, tactile coating surfaces were detected with YOLOv4, YOLOv5, and YOLOv6, which is a version of the deep learning-based object detection method YOLO algorithm.

YOLO is a deep learning-based object recognition algorithm developed by Joseph Redmon. YOLO performs object detection using a single neural network to estimate class probabilities along with bounding boxes. YOLO's architecture is based on a convolutional neural network and has a fast structure [40].

YOLO treats object detection as just a regression problem, using a single neural network. In addition, YOLO also uses a single neural network framework to estimate bounding boxes and class probabilities. While the detection process is performed, the YOLO algorithm estimates the class and coordinates of all objects in the photo by dividing the relevant photo into grids. Thus, object detection is treated as a single regression problem. YOLO can directly calculate location coordinates and classification of objects without intermediaries. It processes photographs at 150 frames per second, thus providing users with real-time object detection. In this way, it has faster prediction power than other methods. This means that in the area where the object will be detected, it passes the object it needs to detect through the network only once and completes the detection process. Considering the literature, it is the fastest general-purpose object detection method [41].

YOLOv4 was a real-time object detection model released in April 2020 and achieved the highest performance on the COCO dataset. It works by dividing the object detection task into two parts; Bounding boxes are used to determine the class of the object. Secondly, the Regression method is used to describe the position of the object through classifiers. The Yolov4 application uses the Darknet framework. Using YOLOv4, many new contributions have been added to the YOLO family alongside previous research contributions, including new features specific to YOLOv4: CSP, WRC, SAT, CmBN, Mish activation, Mosaic data augmentation, DropBlock, CmBN orchestration, and Like CIoU loss. In summary, better object detection network architecture and new data development techniques are used with YOLOv4 [42].

YOLOv5 is a model in the YOLO family of computer vision models. YOLOv5 is widely used to detect objects. YOLOv5 comes in four main versions: small (s), medium (m), large (l), and extra-large (x), each offering increasingly higher accuracy rates. Each variant also takes a different amount of time to train. In the graph, the goal is to create a high-performance object detection model (Y-axis) based on extraction time (X-axis). Preliminary results show that YOLOv5 performs very well compared to other state-of-the-art techniques for this purpose. YOLOv5 variants run faster than Effi-cientDet. As the most accurate YOLOv5 model, YOLOv5x can process photographs much faster with the same level of accuracy as the Effi-cientDet D4 model. Although YOLOv5 derives most of its performance improvements from PyTorch's training routines, its modeling architecture is still close to YOLOv4 [43].

The YOLOv6 model was released by Meituan in June 2022. It gave good results in the COCO dataset comparison. The YOLOv6 model is built on the basis of the YOLO architecture and offers various improvements and new methods over other models of the YOLO family. YOLOv6 is written in PyTorch. YOLOv6 has three major updates: a hardware-friendly spine, neck design, and dedicated head for efficiency, and effective training strategies.

YOLOv6's object detection performance has been shown to be comparable to other CNN-based algorithms, with improvements in both speed and accuracy with incremental versions of the algorithm. In this study, the effect of different image sizes on the performance of YOLOv6 subarchitectures is shown. The data set used for the YOLOv6 model in this study was derived into two versions with dimensions of 416x416 and 640x640 [44].

YOLOv7 has a fast and powerful network architecture. It achieves this with four features it has. These features, integration method, increased label assignment, and model training efficiency. YOLOv7 requires much less computing hardware than other deep learning models. With YOLOv7, it can be trained quickly with small datasets even without pre-trained weights. These features make YOLOv7 a more effective model. Object tracking algorithms generally divide the image into regions, select possible regions containing objects, and classify each region separately, which increases the processing load [45].

YOLOv8 is equipped with more advanced post-processing techniques than its previous versions. These techniques are applied to estimated bounding boxes and object scores produced by the YOLOv8 neural network. With YOLOv8, it refines detection results, removes unnecessary detections, and increases the overall accuracy of predictions. These techniques include Soft-NMS, a variant of the non-maximum suppression (NMS) technique used in YOLO architectures. Instead of completely deleting overlapping bounding boxes, Soft-NMS applies a soft threshold to them. YOLOv8 passes the image through the convolutional ANN at once and performs group normalization. This provides a frame processing speed that can vary from 5 FPS to 160 FPS [46].

3. Results and Discussion

A computer with an AMD Ryzen 1500X 3.5 GHz processor was used in this study. NVIDIA GeForce GXT 1050Ti 4GB GDDR5 was preferred as the graphics card. Memory speed is 16 GB 3000 MHz. Firstly, the YOLOv8 dataset architecture was used. The training period for detecting tactile coating surfaces on the YOLOv8 architecture lasted 4 hours, 49 minutes, and 15 seconds. For tactile-coated surfaces, the highest accuracy rate was 97.2% on the stimulating surface and 98.9% on the guiding surface. The average accuracy rate was 97.9%. Table 1 shows the results occurring in the YOLOv8 architecture.

Table 1. Results of YOLOv8 architecture.

Class	Precision	Recall	F1-Score	mAP@.5	mAP@.5:.95
All	0.979	0.905	0.940	0.977	0.789
Stimulating	0.972	0.914	0.942	0.973	0.719
Guiding	0.989	0.896	0.940	0.980	0.859

The guiding and stimulating surfaces on tactile coating surfaces have different accuracy rates. This is due to differences between tactile coating surfaces. The average performance values of the YOLOv8 architecture were Precision 97.9%, Recall 90.5%, F1-Score 0.940, mAP@.5 0.977, mAP@.5:95: 0.789. Figure 4 shows the confusion matrix implemented in the system.

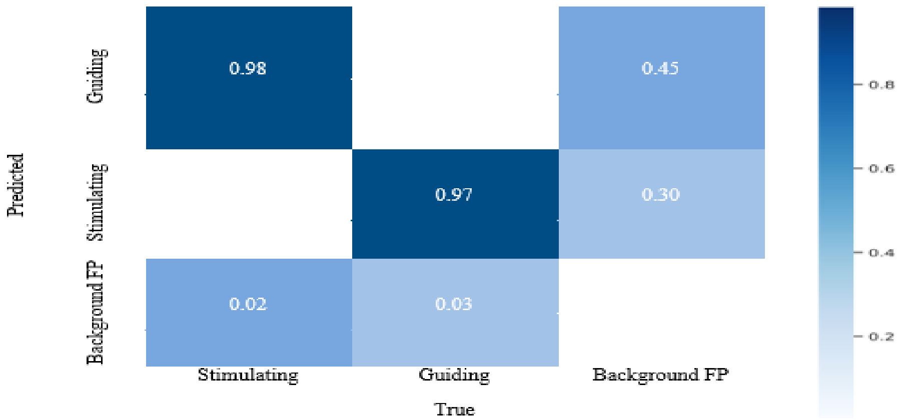


Figure 4. Confusion matrix resulting from YOLOv8 architecture.

A confusion matrix application was carried out to evaluate the performance of this study. The performance of the classification algorithm is visualized with the confusion matrix. For the confusion matrix, 406 images were used for detection. Figure 5 shows the confusion matrix results.

As a result of the confusion matrix, the average accuracy was determined as 95.50%. The confusion matrix results were very close to the simulation results. The confusion matrix results and simulation results show that the study is suitable for the real environment.

In this study, YOLOv5, YOLOv6, and YOLOv7 architectures were used as deep learning. The same data set was used in all studies. Thus, the error rate is minimized. The results for the YOLOv5, YOLOv6, and YOLOv7 architectures used in this study were shown in Figure 6.

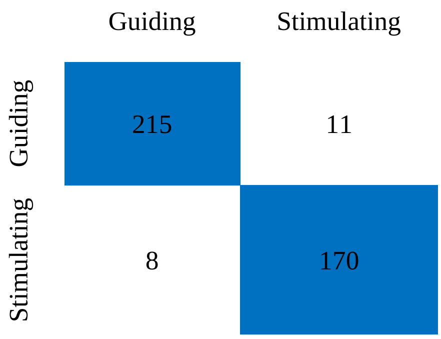


Figure 5. Confusion matrix result.

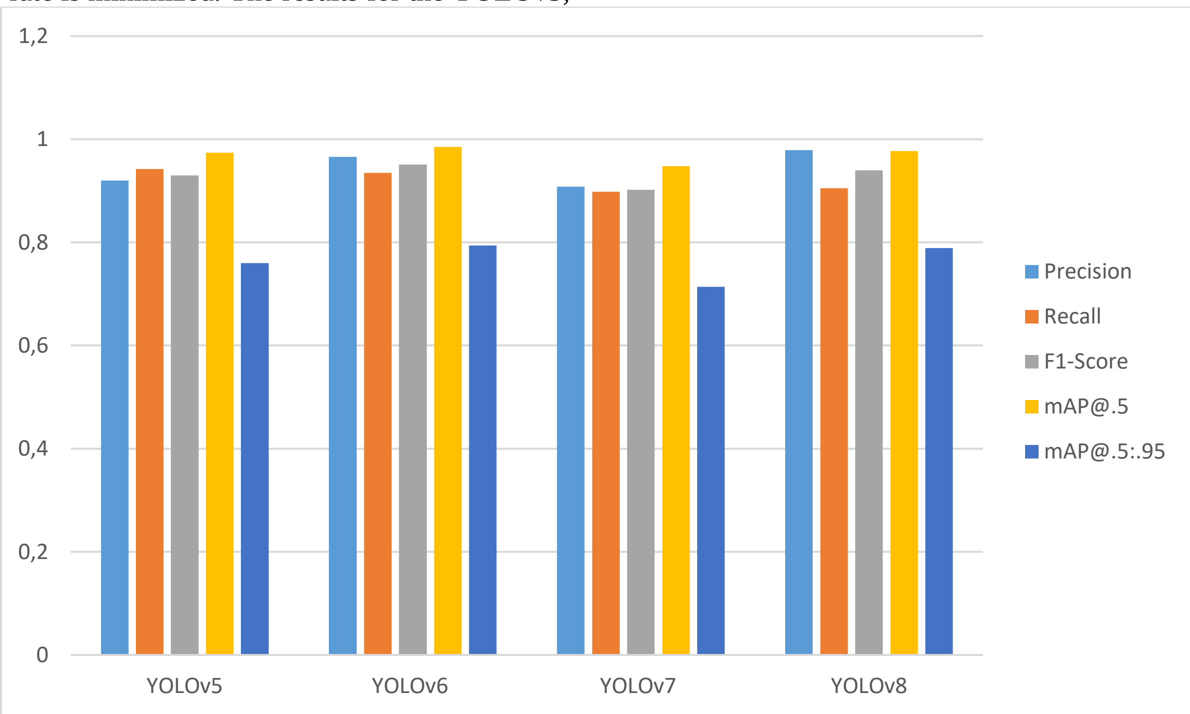


Figure 6. YOLOv5, YOLOv6, and YOLOv7 architectures results.

When the accuracy was examined in this study, the highest accuracy was obtained at 97.9% in the YOLOv8 architecture. The lowest accuracy was 90.8% in the YOLOv7 architecture. The second and third accuracy rates were 92% in the YOLOv5 architecture and 90.8% in the YOLOv7 architecture. When F1 scores are examined, the highest rate occurred in the YOLOv6 architecture. It was later implemented in YOLOv8, YOLOv5, and YOLOv7 architectures.

Since the work will be real-time, the FPS speed must be sufficient. In the YOLOv8 architecture, which gives the highest accuracy rate, the speed has increased up to 150 FPS. At this rate, it is sufficient for real-time detection. Real-time detection was carried out with photographs taken from the camera.

K-Fold Cross Validation method: It divides the data set into “k” equal parts and creates validation data for each part one by one. Thus, each data point is used as validation data at least once. In this way, the overall performance of the model is evaluated more accurately. There are 6278 samples in the data set of the study. Here, the number k is determined as 5. In other words, the data set is divided into 5. There are 1255 samples in each dataset. Their distribution is distributed in the same proportions as in the general data set. Figure shows the results after the K-Fold Cross Validation method.

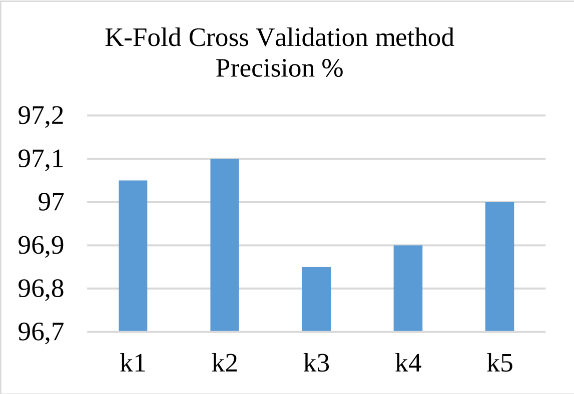


Figure 7. K-Fold Cross Validation method.

The average accuracy rate of the results obtained after the K-Fold Cross Validation method was 97.30. With this result, the overall result of the system is equal to the two results checked. Thus, it is understood that the overall performance of the model is good.

In Figure 8, tactile coating surfaces have been determined in 9 different ways. To ensure that the system adapts to real life, it has been implemented at different times and in different environments.

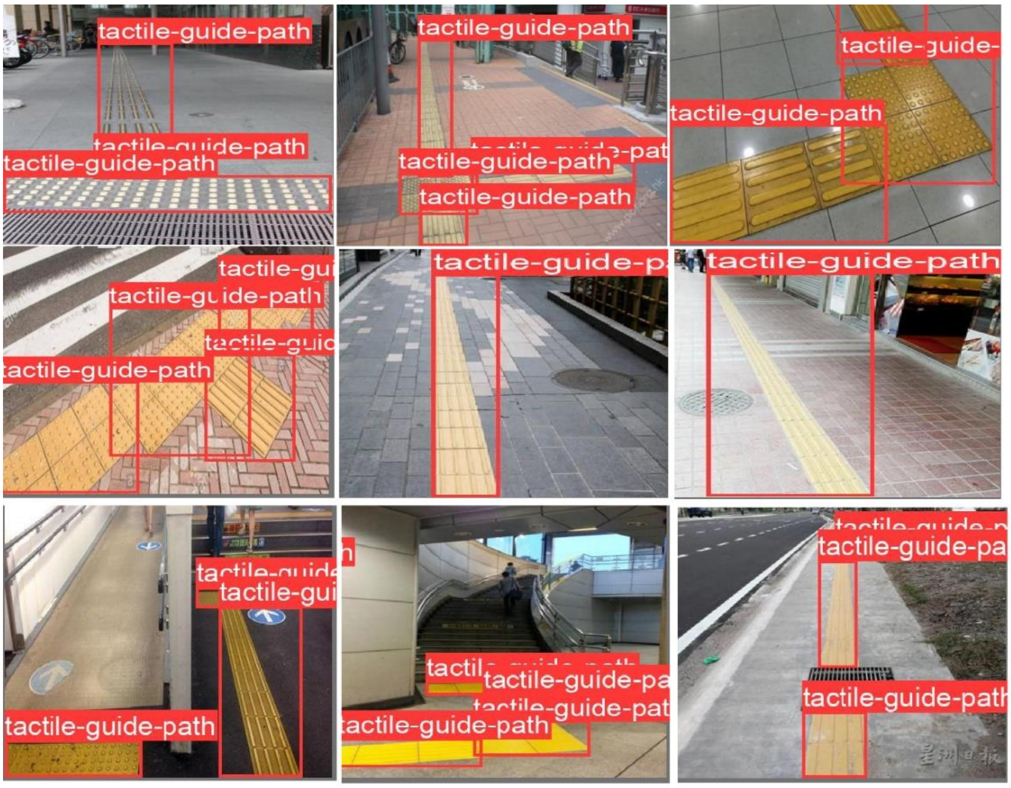


Figure 8. Detection of different tactile-coated surfaces.

When the literature studies are examined, many studies have been done. In these studies, ready-made data sets were not always used. In some studies, the most appropriate data set for the system to be prepared was prepared. Table 2 shows the comparison of literature studies.

Table 2. Comparison of literature studies.

Writer	FPS	Number of Image	Architectural	Architectural
Ghilardi et all [22]	16.27	7000	Faster R-CNN	%88.48
Einloft et all [21]	-	8766	Faster R-CNN	-
Jie et all [23]	2	1586	Faster R-CNN	-
Redmon et all [33]	70	6772	YOLOv3-Tiny	%71
Redmon et all [34]	60	350	Dronet	%80
Asami et all [24]	50	32	YOLOv5	%92
Wang et all [30]	60			%92

As a result of the literature review, it is understood that the usability of previous systems in daily life is quite difficult. These difficulties are that they have a low FPS value. For this reason, the systems made cannot be used in real-time. The systems made have focused only on the tactile coating in a certain place. In this case, different tactile coatings cause the surfaces to be detected at a very low rate or not detected at all.

In this study, four of the latest versions of the YOLO architecture were used. In this way, the highest accuracy rate and FPS speed have been determined by making a comparison. FPS speeds are very good because YOLO architectures take only one look at the image and detect it. The speed of 150 FPS has been reached in the YOLOv8 algorithm. In this way, the detection process can be performed very easily in real-time.

4. Conclusion and Suggestions

This study was not prepared for just one location. It is prepared according to the environments you will encounter in real life. Ready-made data sets were not used for this. A unique data set was prepared for the study. Images from many different environments were used for the data set. Each tactile coating surface in these images is individually labeled. Therefore, it was more easily applied in the real environment.

Additionally, YOLO architectures, the best deep learning method for real-time detection, were used. To achieve the best results in these architectures, YOLOv5, YOLOv6, YOLOv7, and YOLOv8 architectures were used. The highest result was 97% accuracy in the YOLOv8 architecture. A k-fold cross validation process was applied to measure the performance evaluation of the study. Ultimately, it proved the performance of the model. In order for this study to work in real time, the FPS speed must be good. In the study, the FPS rate was increased up to 150 with the YOLOv8 architecture.

Statement of Research and Publication Ethics

The study is complied with research and publication ethics.

References

[1] World Health Organization, Blindness and vision impairment, https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. (30. 01. 2024).

[2] J. Lu, K. Siu, and P. Xu, "A comparative study of tactile paving design standards in different countries," *9th International Conference on Computer-Aided Industrial Design and Conceptual Design*, Kumning, China, 753-758, 22-25 Nov. 2008. https://doi.org/10.1109/CAID/CD15004.2008

[3] A. Mancini, E. Frontoni, and P. Zingaretti, "Mechatronic System to Help Visually Impaired Users During Walking and Running," *IEEE Transactions on Intelligent Transportation Systems*, 2018, 19(2), 649-660. https://doi.org/10.1109/TITS.2017.2780621

[4] J. Redmon, and A. Farhadi, "Yolov3: An incremental improvement," *Computer Vision and Pattern Recognition*, 2018, 1-15. https://doi.org/10.48550/arXiv.1804.02767

[5] G. Huang, Z. Liu, and L. Maaten, "Weinberger, K. Q. Densely Connected Convolutional Networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii-USA, 2261-2269, 21-26 July 2017. https://doi.org/10.5555/2149960

[6] A. Garcia-Garcia, S. Orts, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Rodríguez, "A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation," *Applied Soft Computing*, 2018, 70, 41-65. https://doi.org/10.1016/j.asoc.2018.05.018

[7] A. Krizhevsky, I. E. Sutskever, and G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems*, 2012, 25(2). https://doi.org/10.1145/3065386

[8] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li, "Ieee. ImageNet: A Large-Scale Hierarchical Image Database," *IEEE-Computer-Society Conference on Computer Vision and Pattern Recognition Workshops*, Miami Beach, FL, USA, 248-255, 20-25 Jun. 2009. https://doi.org/10.1109/CVPRWorkshops15504.2009

[9] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, and F. F. Li, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision*, 2015, 115, 211-252. https://doi.org/10.48550/arXiv.1409.0575

[10] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," *Proceedings of the IEEE international conference on computer vision*, Massachusetts, USA, 1026-1034, 7-13 Oct. 2015. https://doi.org/10.1109/ICCV33071.2015

[11] M. Tan, and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," *2019 International Conference on Machine Learning*, California-USA, 6105-6114, 14-16 February 2019. https://doi.org/10.1109/COMITCon45641.2019

[12] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *Proceedings of the IEEE conference on computer vision and pattern recognition*, Columbus, OH, USA, 580-587, 25 Sept. 2014. https://doi.org/10.1109/CVPRW34339.2014

[13] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A.C. Berg, "Ssd: Single shot multibox detector," *European conference on computer vision(ECCV2016*), Amsterdam,Netherlands, 21-37, 11-14 Oct. 2016. https://doi.org/10.1007/978-3-319-46466-4

[14] A. Kaya, A.S. Keçeli, and A.B Can, "Examination of various classification strategies in classification of lung nodule characteristics," *Journal of the Faculty of Engineering and Architecture of Gazi University*, 2019, 34(2), 709-725. https://doi.org/10.17341/gazimmfd.416530

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *IEEE conference on computer vision and pattern recognition*, Las Vegas, NV, USA, 779-788, 27-30 June 2016. https://doi.org/10.1109/CVPR33180.2016

[16] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks", *International Conference on Learning Representations,* Scottsdale, Arizona, USA, 2-4 May 2013. https://doi.org/10.48550/arXiv.1312.6229

[17] Y. Du, N. Pan, Z. Xu. F. Deng, Y. Shen, and H. Kang, "Pavement distress detection and classification based on YOLO network," *International Journal of Pavement Engineering*, 2020, 1659-1672. https://doi.org/10.1080/10298436.2020.1714047

[18] J. Shen, N. Liu, H. Sun, X. Tao, and Q. Li, "Vehicle Detection in Aerial Photographs Based on Hyper Feature Map in Deep Convolutional Network," *KSII Transactions on Internet & Information Systems*, 2019, 13(4).No:4. https://doi.org/10.3837/tiis.2019.04.014

[19] K. Arai, Y. Ando, and M. Mizukawa, "Development of signal recognition improvement system in crosswalk", *Jsme Robomec*, 2008, pp.2A1- E12(1)-(2). https://doi.org/10.1299/jsmermd.2008._2A1-I03_1

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 39,No:6. https://doi.org/10.1109/TPAMI.2012.205

[21] D. C. Einloft, M.C. Ghilardi, and I. H. Manssour, "Automatic Detection of Tactile Paving Surfaces in Indoor Environments", *Workshop of Undergraduate Works (WUW) in the 29th Conference on Graphics, Patterns and Photographs*, Sau Paul, Brazil, 4-7 Oct. 2016. https://doi.org/10.1109/SIBGRAPI.2016.001

[22] M. C. Ghilardi, R.C.O Macedo, and I. H. Manssour, "A New Approach for Automatic Detection of Tactile Paving Surfaces in Sidewalks", *Procedia Computer Science*, 2016, 80, 662-672. https://doi.org/10.1016/j.procs.2016.05.356

[23] X. Jie, W. Xiaochi, and F. Zhigang, "Research and implementation of blind sidewalk detection in portable eta system", *International Forum on Information Technology and Applications*, Kunming, China, 431-434, 16-18 July 2010. https://doi.org/10.1109/IFITA.2010.360

[24] T. Asami, and K. Ohnishi, "Crosswalk location, direction and pedestrian signal state extraction system for assisting the expedition of person with impaired vision", *10th France-Japan/8th Europe-Asia Congress on Mecatronics*, Tokyo, Japan, 285-290, 27- 29 Nov. 2014. https://doi.org/10.1109/MECATRONICS33801.2014

[25] A. Kassim, T. Yasuno, M. S. Mohd, A. Hjshukor, H.I. Jaafar, F. Baharom, and F. Jafar, "Vision based of tactile paving detection method in navigation system for blind person", *Jurnal Teknologi*, 2015, 77, 20. https://doi.org/10.5772/intechopen.79886

[26] O. Yildiz, "Melanoma detection from dermoscopy photographs with deep learning methods: A comprehensive study", *Journal of the Faculty of Engineering and Architecture of Gazi University*, 2019, 34(4), 2241-2260. https://doi.org/10.17341/gazimmfd.435217

[27] E. Dandil, and R. Polattimur, "Dog Behavior Recognition and Tracking based on Faster R-CNN", *Journal of the Faculty of Engineering and Architecture of Gazi University*, 2019, 35(2), 819-834. https://doi.org/10.17341/gazimmfd.541677

[28] J.J. Lv, X.H. Shao, J.S. Huang, X.D. Zhou, and X. Zhou, "Data augmentation for face recognition". *Neurocomputing*, 2017, 230, 184-196. https://doi.org/10.1016/j.neucom.2016.12.025

[29] S. Longfei, Y. Zheng, Z. Zimu, Z. Xiaolong, W. Chenshu, and L. Yunhao, "Crossnavi: Enabling real-time crossroad navigation for the blind with commodity phones", *In Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing, UbiComp 14*, pages 787–798, New York, USA, 1-4 September 2014. https://doi.org/10.1145/2632048.2632083

[30] J. Wang, N. Wang, L. Li, and Z. Ren, "Real-time behavior detection and judgment of egg breeders based on YOLOv3", *Neural Computing and Applications*, 2019, 32, 2. https://doi.org/10.1007/s00521-019-04645-4

[31] W. Byung, Y. Sung, and J. Kang, "Brick path detection from shape pattern and texture feature", *In System Integration (SII), 2011 IEEE/SICE International Symposium on*, Kyoto, Japan, 78–83, 1 Dec 2011. https://doi.org/10.1109/SII.2011.6147423

[32] S. Shoval, J. Borenstein, and Y. Koren, "The NavBelt-a computerized travel aid for the blind based on mobile robotics technology", *IEEE Transactions on Biomedical Engineering*, 1998, 45(11), 1376-1386. https://doi.org/10.1109/10.725334

[33] J. Redmon, and A. Farhadi, "YOLO9000: Better, Faster, Stronger", *2017 IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii-USA, 6517- 6525, 21-26 July 2017. https://doi.org/10.1109/CVPR35066.2017

[34] C. Goutte, and E. Gaussier, "A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation", *Proceedings of the 27th European conference on Advances in Information Retrieval Research, Spain*, 345-359, 21-23 March 2005. https://doi.org/10.5555/2149960

[35] R. Redmon, and A. Farhadi, "Yolov3: An incremental improvement", *Computer Vision and Pattern Recognition*, 2018, 1-6. https://doi.org/10.1007/s11263-009-0275-4

[36] R. Pyun, Y. Kim, P. Wespe, and R. Gassert, "Advanced augmented white cane with obstacle height and distance feedback", *IEEE Int. Conf. Rehabil. Robotics*,Seattle, WA, USA, pp.1-6. 24-26 June 2013. https://doi.org/10.1109/ICORR21102.2013

[37] A. M. Kassim, M. H. Jamaluddin, M. R. Yaacob, N. S. N. Anwar, Z. M. Sani, and A. Noordin, "Design and development of MY 2nd EYE for person with impaired vision", *IEEE Symposium on Computers & Informatics*, Kuala Lumpur, Malaysia, pp.700-703. 20-23 March 2011. https://doi.org/10.1109/ISCI17961.2011

[38] D. Ahmetovic, C. Bernareggi, A. Gerino, and S. Mascetti, "Zebrarecognizer: Efficient and precise localization of pedestrian crossings", *In Pattern Recognition (ICPR), International Conference on Pattern Recognition*, Stockholm, Sweden, 2566–2571, 24-28 Aug 2014. https://doi.org/10.1109/ICPR33632.2014

[39] A. M. Kassim, H .İ. Jaafar, M. A. Azam, N. Abas, and T. Yasuno, "Design and development of navigation system by using rfid technology", *In System Engineering and Technology (ICSET), IEEE International Conference on System Engineering and Technology*, Shah Alam Malaysia, pages 258–262, 19-20 Aug. 2013. https://doi.org/10.1109/ICSEngT.2013.6650181

[40] K. W. M. Jiangyan, and X. Ping, "A comparative study of tactile paving design standards in different countries", *International Conference on Computer-Aided Industrial Design and Conceptual Design*, Kunming, China, pages 753–758, 22-25 Nov. 2008. https://doi.org/10.1109/CAID/CD15004.2008

[41] V. N. Murali, and J. M. Coughlan, "Smartphone-based crosswalk detection and localization for visually impaired pedestrians", I*n Multimedia and Expo Workshops (ICMEW), 2013 IEEE International Conference on*, San Jose, California, USA, pages 1–7, 15-19 July 2013. https://doi.org 10.1109/ICMEW32311.2013

[42] H. Shuihua, P. Hangrong, Z. Chenyang, and T. Yingli, "Rgb-d image-based detection of stairs, pedestrian crosswalks and traffic signs", *Journal of Visual Communication and Image Representation*, 2014, 25(2): 263-272. https://doi.org/10.1016/j.jvcir.2013.11.005

[43] W. Shuihua, and T. Yingli, "Detecting stairs and pedestrian crosswalks for the blind by rgbd camera", *In Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference on*, Philidelphia, USA, 732–739, 4-7 Oct. 2012. https://doi.org/10.1109/BIBMW.2012.6470197

[44] C. Shorten, and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning", *Journal of Big Data*, 2019, 6(1), No: 60. https://doi.org/10.1186/s40537-019-0197-0

[45] C. Y. Wang, A. Bochkovskiy, and H. Y. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new stateof-the-art for real-time object detectors", *Computer Vision and Pattern Recognition*, 2022, 1-15. https://doi.org/10.48550/arXiv.2207.02696

[46] B. Hu, M. Zhu, L. Chen, L. Huang, P. Chen, and M. He, "Tree species identification method based on improved YOLOv7", *IEEE 8th International Conference on Cloud Computing and Intelligent Systems*, 622-627, Chengdu, China, 26-28 November 2022.