Sebastian BUDZAN[1], Mateusz SZWEDKA[1]

Silesian University of Technology, Department of Measurements and Control Systems (1)

doi:10.15199/48.2024.12.12

Integration of YOLO detection algorithm with trajectory prediction of pedestrians for advanced driver assistance system

Abstract. *The article explores the potential of integrating the YOLOv3 detection algorithm with trajectory prediction in ADAS systems. It presents the concept and analyzes the effectiveness of this combination in various driving scenarios. Additionally, it discusses practical implementation aspects and suggests directions for the development of this solution for advanced driver assistance systems.*

Streszczenie. *A artykule zaprezentowano potencjał integracyjny algorytmu detekcji YOLO w wersji 3 z predykcją trajektorii w systemach ADAS. Przedstawia on koncepcję oraz analizę efektywności tego połączenia w różnych warunkach drogowych. Ponadto omawia praktyczne aspekty implementacji i sugeruje kierunki rozwoju tego rozwiązania dla zaawansowanych systemów wspomagania kierowcy.* ***(Integracja algorytmu detekcji YOLO z przewidywaniem trajektorii ruchu pieszego dla zaawansowanego systemu wspomagania kierowcy)***

Keywords: pedestrian detection, trajectory prediction, ADAS, YOLO
Słowa kluczowe: detekcja pieszych, predykcja trajekorii, ADAS, YOLO

Introduction

Contemporary Advanced Driver Assistance Systems (ADAS) face the challenge of integrating various technologies to ensure effective detection and prediction of behaviors on the road. However, the complexity of the road environment and the diversity of traffic participants' behaviors require continuous improvement of these systems. Currently, ADAS systems include many algorithms, mostly based on video information, including systems such as maintaining the driving path using road line detection, obstacle detection, pedestrian detection, road sign detection, intelligent speed assistance system, fatigue detection, and general driver's condition, parking assistance and rear and all-round video camera.

Pedestrian detection systems are based on increasing road safety, thereby increasing the chance of avoiding a collision between a vehicle and a pedestrian. As the statistics of the World Health Organization show in the document "Global status report on road safety 2023", pedestrians accounted for 23% of all fatalities, and drivers 30% [1]. Pedestrian detection in real applications is a complex process and includes not only detecting the human pose, but also tracking movement, detecting orientation, analyzing intentions, and ultimately predicting potential collisions. In the case of advanced solutions, the ADAS system will also take into account scene analysis and parameters related to vehicle movement. Moreover, pedestrian detection is associated with many factors to which the algorithm should be robust, i.e. the appearance of pedestrians, crowds when the human pose is not fully visible, varied distance from the camera, variable background, and weather conditions.

Taking into account all the above observations, the problem of effective detection should be reduced to the precise detection of the human silhouette and then the estimation of the movement trajectory or analysis of the pedestrian's intentions. For this reason, the article presented the results for various scenarios of pedestrian movement, including those crossing a pedestrian crossing or moving along the sidewalk.

In this work, we described pedestrian intention estimation idea and pedestrian pose detection with trajectory prediction methods. We analyzed the effectiveness of a combination of YOLOv3 and a straightforward trajectory prediction algorithm. We analyzed the obtained results using SYNTHIA DataSet for different moving pedestrian scenarios.

Pedestrian intention estimation (PIE)

The main element of pedestrian behavior prediction systems is the ability to assess pedestrian intentions, which we understand as the intention and action taken by a pedestrian on the road. Effectively assessing these intentions is crucial in a driver's ability to respond to potential adverse reactions from a pedestrian. For this reason, the pedestrian's intentions should be clearly defined using information obtained from the analysis of his pose, movement, action taken, or his surroundings and his interaction with the environment.

The PIE methods can be categorized into three main groups [3], i.e. based on the duration of prediction, the type of features involved, and the type of model used. In the first group, short-term and long-term predictions are distinguished. These methods are primarily intended to predict the movement of pedestrians. Short-term methods allow for predicting the next few seconds, while long-term methods aim to estimate its trajectory or target.

Feature-based methods can be based on pedestrian features, contextual features, or a hybrid solution. Pose [4], head orientation [5], trajectory [6], and displacement is important for pedestrian features. Contextual, on the other hand, is based on the analysis of the pedestrian environment, i.e. social interactions, scene information, and ego-vehicle information. In the case of social interactions, situations when a pedestrian is in a crowd are most important. Scene analysis focuses on assessing whether a pedestrian is near pedestrian crossings, intersections, or stops. Hybrid solutions naturally combine both previously mentioned solutions and are used in the most complex scenarios.

An interesting approach is to base it on the model type. In this group of methods, solutions are often based on prediction with all kinds of filters, e.g. Kalman [7] allowing the prediction of pedestrian traffic trajectories, then we are talking about dynamic models or the use of neural networks, then most often the solutions will be data-driven models. In this group, both CNN and Long Short-Term Memory (LSTM) [8, 9, 10], Game-theory-based models [11], and Graph-based networks models [12] are used. Their common feature is that they are based on data that, when used to train the network, can take into account unusual, rare road situations, but require significant computational expenditure and the preparation of a data set, the quality of which will largely determine the effectiveness of detecting road situations involving pedestrians.

Taking into account that the PIE methods should ultimately be designed for usability in ADAS systems, taking into account the above, we decided to use a hybrid solution in our work that does not focus on specific road situations but is based on a precise human pose detection algorithm along with trajectory estimation. Effective detection of the human pose is crucial due to the possibility of using simpler and faster trajectory prediction methods, but it is also the starting point for the detection of specific road situations involving pedestrians. This approach will naturally be a cascade solution in which the human pose is first detected, and then, based on data about the position of the pose in time, we estimate the pedestrian's trajectory. In this regard, we also conducted research using several selected, typical extreme road situations related to pedestrian movement, i.e. crossing lanes in front of the vehicle, moving on the sidewalk approaching and away from the vehicle.

Pedestrian pose detection and trajectory prediction

Human pose detection algorithms are most often an adaptation of algorithms used for this detection in outdoor conditions or previously used in the case of urban surveillance cameras. The scene seen from the perspective of the car's camera is completely different, and the speed of the car causes the background to change much faster compared to situations in which the camera is stationary and only the person moves. The most commonly used detection methods use the classical approach or neural networks. In the classical approach, the method comes down to extracting features specific to the human figure. The most basic approach is to use the Histogram of Oriented Gradients descriptor most often in combination with the Support Vector Machines classifier [13], which focuses on low-level features, through the information on the intensity of the gradient for the featured nine directions. However, the method has some disadvantages, the main of which is the problem of covering a fragment of the silhouette by another or a completely different object, as well as a high percentage of objects incorrectly classified as human poses, especially for objects whose gradient system is similar to the human pose, e.g. lanterns, power poles, high windows. Among the classical methods using feature extraction, Haar Classifier, and Aggregate Channel Features are also used. These types of methods often use additional operations to improve the efficiency of their algorithms, such as background reduction, filtration, and conversion to other color spaces. On the other hand, a common approach is to use a hybrid approach in which pose detection is combined with the detection of the head [14] or other parts of the human body, but in this case, the authors most often use the second group of methods, i.e. neural networks [15, 16].

These methods most often use Convolutional Neural Networks (CNN), YOLO, Single Shot Detector (SSD), Region-Based Convolutional Neural Networks (R-CNN), Fast R-CNN or Faster R-CNN. The next proposed variants of neural networks are aimed primarily at improving efficiency while reducing computational costs, but the most important reason is the need to take into account the dynamics of the scene in front of the vehicle, the unusual behavior of the pedestrian, as well as the speed of the vehicle and the pedestrian. Importantly, the methods strive not only to increase true positive (TP) detection but to reduce false positive (FP), a problem that is present in human pose detection algorithms.

Detection and prediction of pedestrian trajectories is one of the most important elements of the PIE system. Due to the complexity of the scene in urban conditions, the influence of external factors such as social relationships, scene dynamic, and interactions with other people and other objects in the scene [17] its precise determination in a short time horizon of 5-10 seconds is a key aspect of the effectiveness of the system for predicting pedestrian behavior on the road. The effectiveness of trajectory estimation depends largely on the precision of silhouette detection and also affects its prediction. A popular approach to the problem of prediction is the use of statistical methods, filters, e.g. Kalman [18], advanced tools such as Long-Short Term Memory Networks (LSTM) [19], or knowledge-based approaches [20].

Dataset used for research

During our work, we relied on a public data set, which is recognized as SYNTHIA DataSet [21]. It contains about 100GB of images and is also divided into training and test collections if we want to provide a model based on neural networks. Moreover, this data set has also a pack of detailed information about those images, like the position in the image in both two dimensions and three dimensions. The main advantage is the possibility to provide multiple scenarios thanks to the uniqueness of scenes, e.g.: walking through the road crossing or change of trajectory during the walk on the sidewalk. It is a very wide set (*tabela 3.1*), which contains 13 object classes. [22].

Class	ClassID	Red	Green	Blue
Road	3	128	64	128
Car	8	64	0	128
Sign	9	192	128	128
Pedestrian	10	64	64	0
Cyclist	11	0	128	192

Table 1. Partial table with description of chosen class with its channel intensities

According to license limitations, authors allow the scientists to use their set for research or educational purposes.

Object class	Angle (°)	Two-dimensional format limiting [xmin, ymin, xmax, ymax] (px)
Pedestrian	3.08	[82.16, 170.8, 159.54, 431.45]

Table 2. Table, which contains class etiquette and information about the position

Fig. 1. Pedestrian walking through a road crossing - *000391.png*

To research further detection and prediction information gained from the third column of the second table (Table 2.) will be used. Our work will be focused on implementing proper mathematical operations connected with the YOLOv3 algorithm and detecting pedestrians during their movement (see Fig. 1).

The integration of algorithm

The name YOLO is an acronym for the network, whose full name is *"You Only Look Once"* (see Fig. 2). It uses a single deep convolutional network for both classification and detection. There are some advantages to using it, including that the method takes in the whole image in the learning process, so it is thus possible to spot a class object together with the whole background. The algorithm divides the image into a grid of size SxS. When the focal point falls into one of the grid cells, this grid cell will be responsible for object detection [23]. Its CNN backbone is Darknet53, which was originally a deep convolutional network of 53 layers. A further 53 layers are added for detection purposes. The network also provides for ten times as many cubes as the second version, which unfortunately has a significant impact on speed [24].

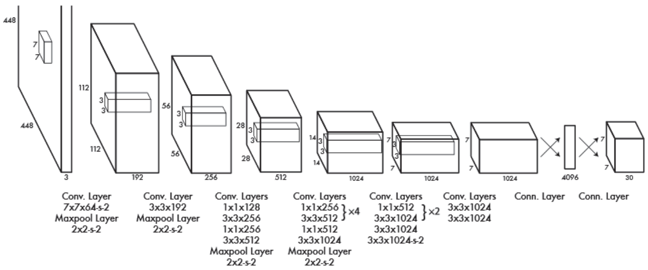
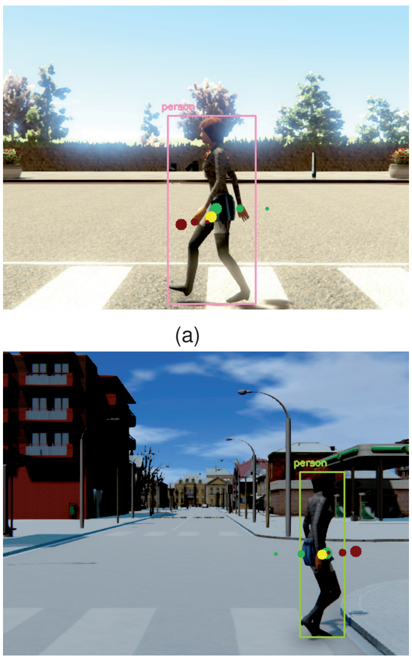


Fig. 2. YOLO's network architecture [25]

Nevertheless, it is not possible to provide a prediction using YOLOv3 alone. For this to be possible, further solutions need to be implemented, and we aimed to determine how complicated it is to provide something comparable in any way to the current ADAS systems. We also decided to test our results in several scenarios in which a pedestrian crosses the road on the left or right, and a pedestrian walks on the pavement in our direction and the opposite direction.

The first scenario is presented below, where we can see pedestrians walking from the right side of the pavement to the left and vice versa.

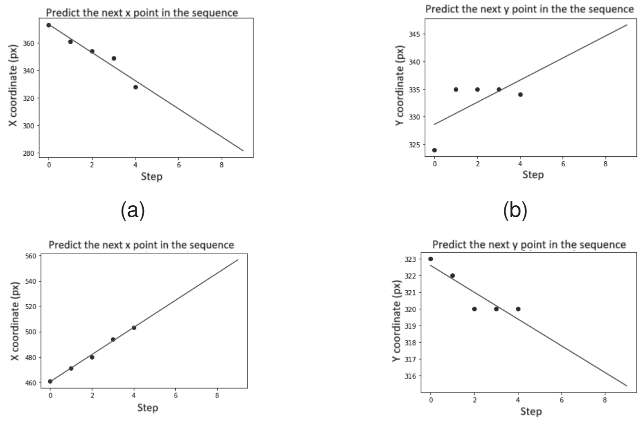


(a)

(b)

Fig. 3. First scenario - pedestrian walking through the road, right to left (a),left to right (b)

We used the data information as a rectangle with a marked human pose calculated by the YOLOv3 algorithm when detecting the desired class (see Fig.3). The rectangle was split in half both vertically and horizontally to get the location of the dot which is crucial for prediction. Using these dot samples, a straightforward algorithm was developed that took five previous samples to predict five future samples. The algorithm was based on tables that generally stored 11 samples. The sixth - last one; from the past was always ommited.



(a) (b)

(c) (d)

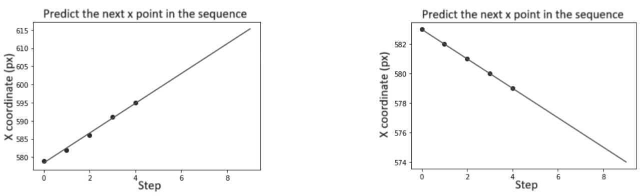
Fig. 4. Prediction results for first scenario – pedestrian walking through the road

As can be seen in Figure 4, the algorithm seems to be very stable - especially in the vertical dimension, where can be observed daily pedestrian movement; and gives promising results for both situations. Figure 4a and Figure 4b are related to the *a* image from Figure 3, and b is related to the *b* image from Figure 3). It was possible to find further points where the pedestrian was supposed to move.

(a) (b)

Fig. 5. Second and third scenario – pedestrian moves toward us (a) and away from us (b)

However, we also presented two other scenarios. These provide an answer to the question of whether the algorithm is "good enough" to provide correct prediction in situations where the pedestrian is walking towards the camera as well as in the opposite direction (see Figure 5). It seems that the results will not be as good for these specific situations and, according to the data presented in the graphs, are only slightly better for the first case.



(a) (b)

Fig. 6. Prediction results for second and third scenario – pedestrian moves toward us (a) and away from us (b)

The prediction results of the second scenario are presented in Figure 6 (a), and the third scenario in Figure 6 (b). Importantly, the differences between the future points estimated from the trajectory prediction are not large enough to effectively determine the direction of pedestrian movement, which is caused by the problem of the proportion of the silhouette size to the image size, often emphasized in the literature, and practice, the distance of the person from the vehicle's camera. On the other hand, the obtained results suggest the correct operation of the human figure detector, but at the same time, despite the correct trajectory prediction, it is not possible to determine the direction of movement.

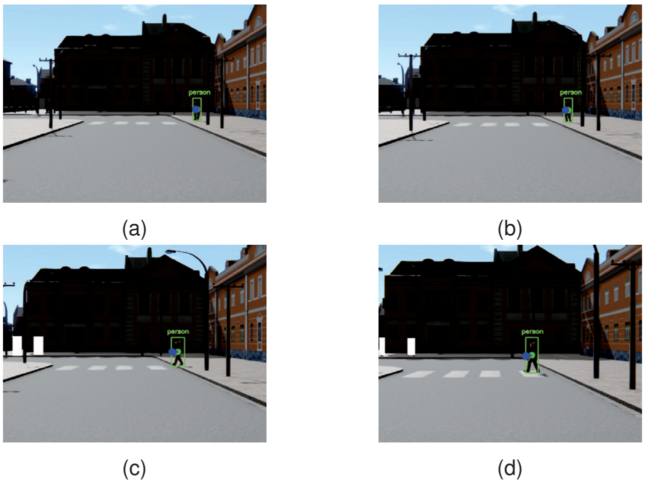


Fig. 7. Example of overlapping scenarios

The situation in which a pedestrian moves towards us and changes direction (see Figure 7, where a, b, c, and d are consecutive captured frames of the same video) to cross the road gives promising results and, according to the graphs in Figure 8 (a, b, c and d are related to the four images in Figure 7), the changes are implemented very quickly, which led us to conclude that this algorithm is a good start to provide much more advanced research to establish better and more precise solutions.

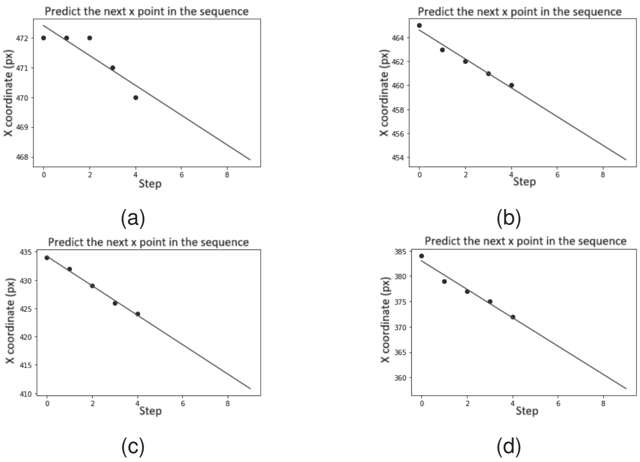


Fig. 8. Prediction results for overlapping scenarios

Unfortunately, as we expected from previous experiments, the results in the scenario where the pedestrian moves away from us and changes trajectory to cross the road (see Figure 9) gave worse results than in the previous multiple scenario case. Nevertheless, the algorithm reacts very quickly to changes which proves its sensitivity in a short time horizon.

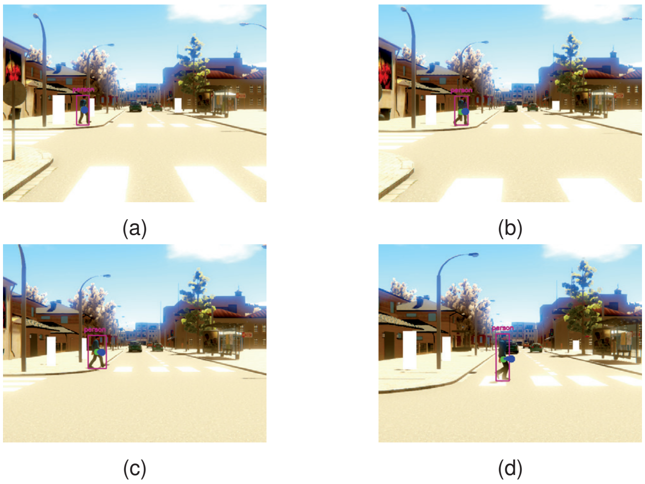


Fig. 9. Scenarios overlap - further observations

We have observed that even in situations where the prediction is not very precise, we are still able to record changes in direction very quickly, and the graphs seem to confirm this (see Figure 10).

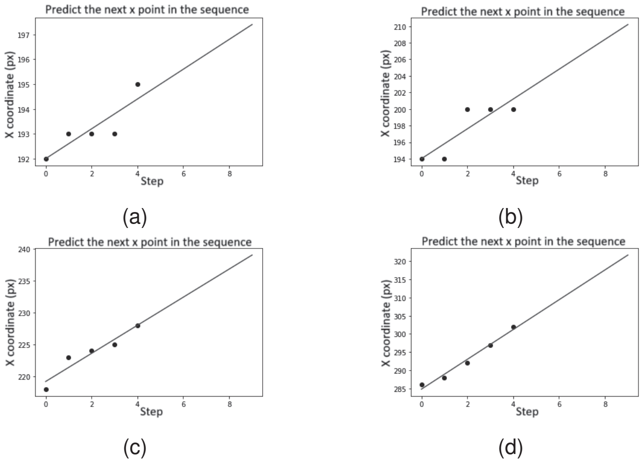


Fig. 10. Scenarios overlap plots - further observations

Determining efficiency

During our work, we have decided to use a few statistical methods to determine if our solution gives any proper results. Those methods are described below.

Pearson correlation coefficient

It allows us to measure the linear relationship between two sets of data. The correlation coefficient varies between -1 and 1, with correlations of -1 or 1 indicating an exact linear relationship, while a value of 0 indicates no correlation. A positive correlation provides information that as the argument x increases, the value of y increases, while a negative correlation does the opposite, i.e. that an increase in the argument x will cause the value of y to decrease [26].

$$\rho = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \tag{1}$$

Estimation however is determined with *equation 2*.

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2 (y_i - \overline{y})^2}} \tag{2}$$

$\overline{x}$ value is mean of x vector, $\overline{y}$ value is mean of y vector.

Spearman's rank correlation coefficient

The Spearman rank correlation coefficient is a non-parametric measure of the monotonicity of the relationship between two data sets. Contrary to Pearson's correlation, Spearman's correlation does not assume a priori that the two data sets have a normal distribution. It also varies between -1 or +1, with 0 indicating no correlation, while correlations of -1 or 1 indicate an exact monotonic relationship. A positive correlation provides the information that as the argument x increases, the value of y increases, while a negative correlation does the opposite, i.e. that an increase in the argument x will cause the value of y to decrease [27].

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{3}$$

The value of *d* represents the distance between pairs x_i and y_i, while *n* represents the number of samples provided.

Kendall's Tau

Kendall's Tau is a measure of concordance between two rankings. Values close to 1 indicate strong agreement and values close to -1 indicate strong disagreement. It measures the degree of monotonicity between variables and, like the Pearson model, calculates relationships between ranked variables, making it applicable to data with a non-normal distribution [28]

$$\tau = \frac{c-d}{c+d} = \frac{S}{\binom{n}{2}} = \frac{2S}{n(n-1)} \tag{4}$$

The value of *c* in *alignment 4* indicates the number of congruent pairs, while *d* indicates the number of incongruent pairs.

Student's "t" Test

This test is defined as a statistical hypothesis test. Its purpose is to test whether two samples were drawn from the same population as expected. It works by testing the means of two samples of data to see if they are significantly different from each other. It does this by calculating the standard error of the difference between the means, allowing itself to be interpreted in such a way that it is possible to see how likely the difference is if the two samples get the same mean. The t-value, on the other hand, can be interpreted by comparing it to outliers obtained from the t-distribution. The outlier itself can be calculated by using the degrees of freedom and the significance level using the percentage point function (FPP). This test is divided into two versions, one is to test unbound samples, i.e. samples that are independent of each other, and the other is to test bound samples, i.e. samples where the samples are dependent on each other, e.g. repeated acquisition of points from an image on the same object. In the case of the second method, the following formula should be used [29].

$$t = \frac{\overline{X_D} - \mu_0}{s_D/\sqrt{n}} \tag{5}$$

In *equation 6*, $\overline{X_D}$ is the mean, while s_D is the standard deviation between all pairs. The constant $micro_0$ is equal to zero when testing whether the mean of the difference is significantly different. The degrees of freedom are equal to *n-1*, where *n* represents the number of samples. By degrees of freedom, it meant the sum of all observations in the two samples minus two.

Results of efficiency determination

All classes used in the study were analyzed, thus different results were obtained. Table 3 containing the results for each is included below. Tests were carried out based on points detected and points predicted on the x axis, as significant changes in direction are observed on this axis only.

Class	Pearson	Spearman	Kendall
1	99%	99%	99%
2	98%	99%	99%
3	70%	71%	63%

Table 3. Correlation between points detected and predictions on the x-axis

Each class hinted at different conclusions, the first of which allows us to conclude that the predicted points are directly correlated with the preceding values and this goes without saying. This confirms that there is a strong, almost 100% correlation between what is detected by the algorithm and what will be predicted by the algorithm. Thus, the algorithm, when a pedestrian is crossing across the roadway, is very accurate and can be used in such a case without problems. This was confirmed by all three tests for both images of this class. When it comes to class two, the results from the table for the tests performed by the Spearman and Kendall models also left no doubt as to the association of these two variables with each other. It is worth mentioning that the received results for the first image of this class were also almost identical. For the third class, the situation became interesting for all correlation coefficients. This unfortunately gave less precise predictions compared to the 2 previous classes. The conclusion that emerges is that camera angles and distances need to be considered, especially for this class, as thin points can overlap significantly. A solution may be to increase the number of steps in the sequence or to create a condition that determines when data is collected for the list.

Student's test results

This subsection uses the Student's test, which was conducted on the same data obtained from the second image of each class. This will lend credence to the validity of the previous results while allowing an additional test to be performed. The data are dependent datasets and will be considered as such. Again, all results have been placed in Table 4 and analysis has been carried out based on this.

Class	Student's T test result
1	98.5
2	42.43
3	10.75

Table 4. Student's test performed on points lying on the x-axis

For the first class high value obtained indicates that the groups analyzed are decidedly different from each other. This is confirmed by the observation and comparison with each other of the data contained in the list containing the detected focal points and the data contained in the predictions. Furthermore, it can be seen that each successive point has a significantly different value from the previous one which facilitates the prediction based on the 5 points. When it comes to the second class it is possible to spot that after analyzing the data contained in both sets, it is undoubtedly noticeable that there is considerably less variation among the successive centers detected, which often differ by only a few pixels. Despite the lack of variation, the results seemed to be acceptable, however, the problem with the appropriate positioning of the camera angle is confirmed here, as the points overlap significantly. The third class, as expected, scored the lowest, this means that more meaningful ways of scoring need to be applied in the case of this class. The differences between the two seem to be symbolic, so predicting trajectories far becomes very problematic.

Conclusion

In our article, we presented the possibility of implementing these solutions in detection algorithms, which could potentially be very useful in ADAS systems. While the application of the algorithm proved to be as possible and relatively simple to implement, the results are unfortunately not highly satisfactory. Under conditions where the pedestrian is observed in cross traffic, the results can be considered sufficient, but the second and third classes defined in this paper, show some shortcomings. We also claim that the possible limitation of this algorithm is its scalability. We have implemented it to detect and predict the trajectory of one person only. If we want to achieve results for a group of pedestrians, we first have to detect and make a collection of central points for each one of them. Nevertheless, it was possible to confirm here the possibility of extending the pre-prepared algorithms with the computing-optimized, author's solutions related to traffic trajectory prediction.

***Authors**: PhD, DSc, Assoc. Prof. Sebastian Budzan, Department of Measurements and Control Systems, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, email: sebastian.budzan@polsl.pl, MSc. Mateusz Szwedka, Department of Measurements and Control Systems, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, email: mateusz.szwedka@gmail.com*

REFERENCES

[1] Global status report on road safety 2023 `https://www.who.int/publications/i/item//9789240086517/`, [Accessed on 27.06.2024]

[2] Roja Ezzati Amini, Kui Yang, Constantinos Antoniou: Development of a conflict risk evaluation model to assess pedestrian safety in interaction with vehicles, Accident Analysis & Prevention, 175 106773, 2022.

[3] Sharma N., Dhiman C., Indu S.: Pedestrian intention prediction for autonomous vehicles: A comprehensive survey, Neurocomputing, 508 pp. 120–152, 2022.

[4] Rasouli A., Kotseruba I., Tsotsos J.K.: Are They Going to Cross? A Benchmark Dataset and Baseline for Pedestrian Crosswalk Behavior, 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), Venice, Italy, pp. 206–213, 2017.

[5] Fang Z., López A.M.: Is the Pedestrian going to Cross? Answering by 2D Pose Estimation, 2018 IEEE Intelligent Vehicles Symposium (IV), Changshu, China, pp. 1271–1276, 2017.

[6] Wu H., Wang L., Zheng S., Xu Q. Wang J.: Crossing-Road Pedestrian Trajectory Prediction Based on Intention and Behavior Identification, 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, pp. 1–16, 2020.

[7] Sung K.: Pedestrian Positioning Using an Enhanced Ensemble Transform Kalman Filter, Sensors, 23(15):6870, 2023.

[8] Marginean, A.; Brehar, R.; Negru, M.: Understanding pedestrian behaviour with pose estimation and recurrent networks, In Proceedings of the 2019 6th International Symposium on Electrical and Electronics Engineering (ISEEE), Galati, Romania, pp 1–6, 2019.

[9] Lorenzo J., Parra I., Wirth F., Stiller C., Llorca D.F., Sotelo M.A.: RNN-based Pedestrian Crossing Prediction using Activity and Pose-related Features, 2020 IEEE Intelligent Vehicles Symposium (IV), Las Vegas, NV, USA, pp 1801–1806, 2020.

[10] Gesnouin J., Pechberti S., Bresson G., Stanciulescu B., Moutarde F.: Predicting intentions of pedestrians from 2d skeletal pose sequences with a representation–focused multi–branch deep learning network, Algorithms, 13(12) pp. 1–23, 2020.

[11] Ezzati Amini R., Dhamaniya A., Antoniou C.: Towards a Game Theoretic Approach to Model Pedestrian Road Crossings, Transportation Research Procedia, 52 pp 692–699, 2021.

[12] Song X., Kang M., Zhou S., Wang J., Mao Y., Zheng N.: Pedestrian Intention Prediction Based on Traffic-Aware Scene Graph Model, 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Kyoto, Japan, pp 9851–9858, 2022.

[13] Surasak T., Takahiro I., Cheng C. -h., Wang C. -e., Sheng P. -y.: Histogram of oriented gradients for human detection in video, 5th International Conference on Business and Industrial Research (ICBIR), Bangkok, Thailand, pp 172–176, 2018.

[14] Saqib M., Khan S. D., Sharma N., Blumenstein M.: Person Head Detection in Multiple Scales Using Deep Convolutional Neural Networks, 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, Brazil, pp 1–7, 2018.

[15] Zhang Y., Lin J.: Research on pedestrian occlusion detection based on SSD algorithm, In Proceedings of the 2019 International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI '19). Association for Computing Machinery, New York, NY, USA, pp 417–421, 2019.

[16] Wu Y., Chen C. Wang B.: Pedestrian Detection Based on Improved SSD Object Detection Algorithm, 2022 International Conference on Networking and Network Applications (NaNA), Urumqi, China, pp 550–555, 2022.

[17] Zhong J., Sun H., Cao W., He Z.: Pedestrian Motion Trajectory Prediction With Stereo-Based 3D Deep Pose Estimation and Trajectory Learning, IEEE Access, 8, pp 23480–23486, 2020.

[18] Lin, C.-Y.; Kau, L.-J.; Chan, C.-Y: Bimodal Extended Kalman Filter-Based Pedestrian Trajectory Prediction, Sensors, 22(8231), 2022.

[19] Quan R., Zhu L., Wu Y., Yang Y.: Holistic LSTM for Pedestrian Trajectory Prediction, IEEE Transactions on Image Processing, 30, pp 3229–3239, 2022.

[20] Korbmacher R., Tordeux A.: Review of Pedestrian Trajectory Prediction Methods: Comparing Deep Learning and Knowledge-Based Approaches, IEEE Transactions on Intelligent Transportation Systems, 23(12), pp 24126–24144, 2022.

[21] Bengar, J.Z., Gonzalez-Garcia, A., Villalonga, G., Raducanu, B., Aghdam, H.H., Mozerov, M.G., López, A.M., Weijer, J.V.: Temporal Coherence for Active Learning in Videos, 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 914–923, 2019.

[22] *SYNTHIA: The SYNTHetic collection of Imagery and Annotations*, Universitat Autònoma de Barcelona, [web page] `http://synthia-dataset.net/table-classes/`, [Accessed on 29.05.2024]

[23] Xupeng Kou, Shuaijun Liu, Kaiqiang Cheng, Ye Qian Development of a YOLO-V3-based model for detecting defects on steel strip surface Measurement, Volume 182, 3pp, September 2021, 109454

[24] Redmon J., Farhadi A., YOLOv3: An Incremental Improvement, Washington University, 2018

[25] Redmon J., Divvala S., Girschick R., Farhadi A., You Only Look Once: Unified, Real-Time Object Detection University of Washington, 2016

[26] Pearson Correlation and Linear Regression `http://sites.utexas.edu/sos/guided/inferential/numeric/bivariate/cor/`, [Accessed on 06.06.2024]

[27] Zwillinger D., Kokoska S., CRC Standard Probability and Statistics Tables and Formulae. Chapman and Hall: New York., 2000, Section 14.7

[28] Maurice G. Kendall, Rank Correlation Methods (4th Edition), Charles Griffin and Co., 1970.

[29] Timothy C. Urdan, Statistics in Plain English - 4th Edition, Routledge, New York, 2016