Using Image Processing for Autonomous Illumination Sensor Optimization

Paween Pongsomboon
Master of Information Technology
Frankfurt University of Applied Sciences
Matriculation Number: 1346048
paween.pongsomboon@stud.fra-uas.de

***Abstract*—This project deals with the development of program to detect person from overhead camera. The purpose of the project is to collect prediction label of data between person and non-person situation. Data and label can then be used to optimize the autonomous illumination sensor. Well-known classification model YOLOv5 was used to classify person/non-person situation. The developed program can work in real-time mode, support reading from video file, and can record and export the prediction result with timestamps. Moreover the program can handle region of interest specification and predict result accordingly. This paper mainly focuses on the development and validation of person detection program and data format. Prediction label exported from the program can be further used with the optimisation process.**

***Index Terms*—Machine Learning, Optimization, Threshold tuning, Data classification, Classification, Image processing, Convolutional Neural Network, YOLOv5**

I. Introduction

As the hardware performance accelerated in the first decade of this century, digital transformation dramatically improved. As a result, automation is becoming common in daily life. We may all experience autonomous illumination at some point in our daily lives. Many autonomous lighting systems use light sensors or motion sensors as their primary sensor, but this paper would like to introduce an alternative: ultrasonic sensors. For an autonomous illumination system to be successful, consistency is a key component, and ultrasonic sensors offer a lot of room for improvement. It is essential that the system be capable of accurately detecting human presence.

A. Human detection and image processing

Human detection systems have rapidly developed in the past decade, courtesy of the development of hardware performance and many new inventions in machine learning, Convolutional Neural Network (CNN), which is inspired by the organization of animal visual cortex [4] [5]. CNN opened the floodgate to the development of object classification in image processing field, which is an important key in learning mechanism. Rikiya Yamashita et al. explained CNN as a mathematical construct that is typically composed of three types of layers : convolution, pooling, and fully connected layers. These three types of layers play important roles in every CNN models nowadays. The Neocognitron [5] developed by Kunihiko Fukushima has led to the development of many adaptive CNN models. In 2015, Joseph Redmon et al. developed one of the most popular object detector CNN called YOLO (You Only Look Once), which their works was motivated by Region Convolutional neural network (R-CNN) [1].

B. YOLOv5

Joseph Redmon et al. published the first of YOLO invention series in 2015, which used the conceptual design of the bounding boxes to predict the object with score. Each bounding boxes will have its own prediction class and confidence score. The neural network of YOLOv1 is inspired by the GoogLeNet model for image classification [1]. The conceptual design of YOLO is very efficient and lead to the further development until version 5 (YOLOv5), which is used in this person detection in this paper. Yolov5 is the 5th edition of its famous YOLO series, which is known for its accuracy and speed [2]. Glenn Jocher introduced YOLOv5 using the Pytorch framework shortly after the publication of YOLOv4 by Alexey Bochoknovskiy et al [3] in 2020. YOLOv5 pretrained with the Microsoft Common Objects in Context (COCO) dataset and has many application programming interface (API) available in python programming language, which has the ease of access to many developers.

C. Working Principle of Ultrasonic Sensor

Ultrasonic Sensors work on the principle of Echolocation. This is the principle used by bats to maneuver without colliding into obstacles. Sound waves are emitted and based on the time, it takes for these waves to hit an object and return, the position of the object is calculated, as shown in (Fig. 1) . Since the speed of sound in air, v_{sound}, is assumed to be 343 m/s, the following equation (1) gives the distance of the object, d, from the sensor :

$$d = \frac{(t_{roundtrip} \times v_{sound})}{2} \tag{1}$$

where $t_{roundtrip}$ is the total time for the wave to hit the object and travel back to the sensor and the value is divided by 2 to only get the one way trip distance. This time-of-flight measurement with sound waves above 20 kHz gives accurate results. This precision combined with the below-listed advantages make ultrasonic sensor a viable, reliable and practical option for person detection.

D. Advantages of Ultrasonic Sensor over Others

- IR sensors need knowledge of the surface of reflection and can be affected by dust, smoke, fog and the like, that blocks line of vision. But Ultrasonic sensors are not affected by any of these.

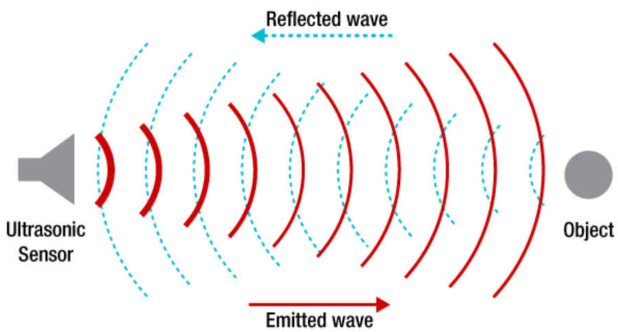

Fig. 1. Ultrasonic Time-of-Flight Measurement

- Optical based sensing technologies do not work well in bright ambient lighting, smoky or foggy conditions and cannot detect water nor glass. Whereas ultrasonic waves bounce off glass and water, detection is possible
- RADAR and LIDAR solve the above mentioned problems and function with great accuracy, but are too expensive

Above all, Ultrasonic sensors are much more affordable and have already been employed in driver and parking assistance systems. Adding another sensor to this array is much more achievable and practical.

II. System Architecture

This section describes the overall system and how the experiment was setup. The system architecture, in this experiment, consists of two subsystems, the Data Labeling system and Ultrasonic Data Collection as illustrated in Fig. 2. The Data Labeling system deals with the person detection in the area of the experiment, while the Ultrasonic Data Collection system collects data from ultrasonic sensor at the same time. As a result, the CNN model can be fed data from ultrasonic sensors and labels in optimization.

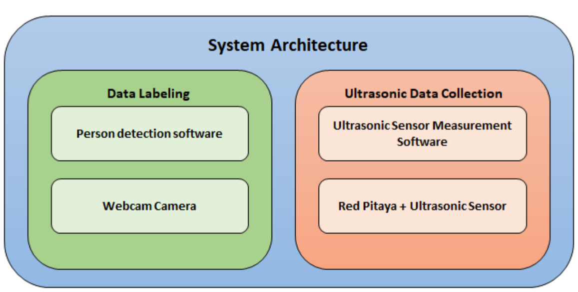

Fig. 2. System Architecture

A. Experimental Setup

Fig. 3 illustrates the side view of the experimental setup, which shows components from two sub systems. The Data Labeling system consists of the camera attached to the ceiling of the room and a person detection software inside a computer, while the Ultrasonic data Collecting system consists of the ultrasonic sensor, which is mounted on top of Red Pitaya device, and the Ultrasonic Sensor Measurement software. The camera is attached to the ceiling of the room, therefore it can cover the region of interest, which is the area covered by ultrasonic beam. The camera is The Ultrasonic sensor is mounted on the structure, the center of sonar is 203.30 cm high from the floor, while the desk is 77.80 cm high from the floor. Triangular shapes in Fig. 3 represents the coverage area of ultrasonic beam, which can be calculated simply from Pythagoras's equation. The coverage area of ultrasonic beam is a circular area with radius of 94.8 cm on the floor and radius of 58.52 on the desk as illustrated in Fig. 4. Moreover the experimental setup in real place is illustrated in Fig. 5 and the coverage area from overhead view is illustrated in Fig. 6.

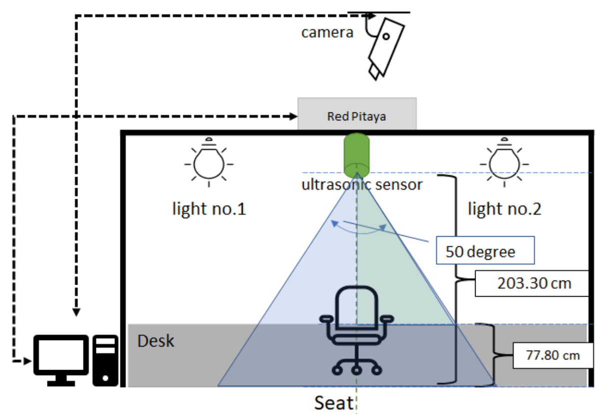

Fig. 3. Experimental Setup Diagram

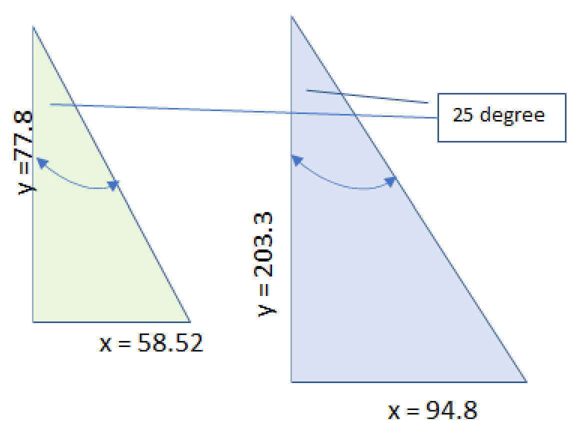

Fig. 4. Ultrasonic beam radius

Fig. 5. Experimental Setup in real place

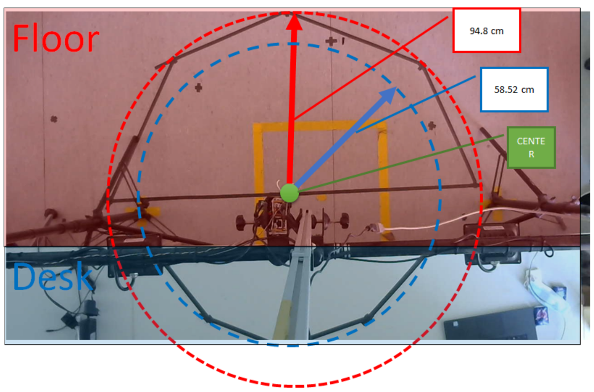

Fig. 6. Overhead view of experiment coverage area

B. Data Labeling

Data labeling plays the crucial part in this paper, since it provides the labels for data from the ultrasonic sensor. Without a doubt, image processing method provides more robustness and accuracy for person detection, therefore, it is selected as a data label tool for optimisation. In order to do so, the person detection program was developed and evaluated.

The person detection program uses a camera to capture the scene and then processes frame by frame to detect a person. The software, which is illustrated in Fig. 7, connects to the camera and can detect a person in the region of interest (ROI).

1) Development of the software: The person detection software was developed using python programming language and its source code can be found in [8]. The software used the YOLOv5 library from Pytorch as an object detector. The software supports three modes of detection, which are real-time mode, video mode, and record more. The software also

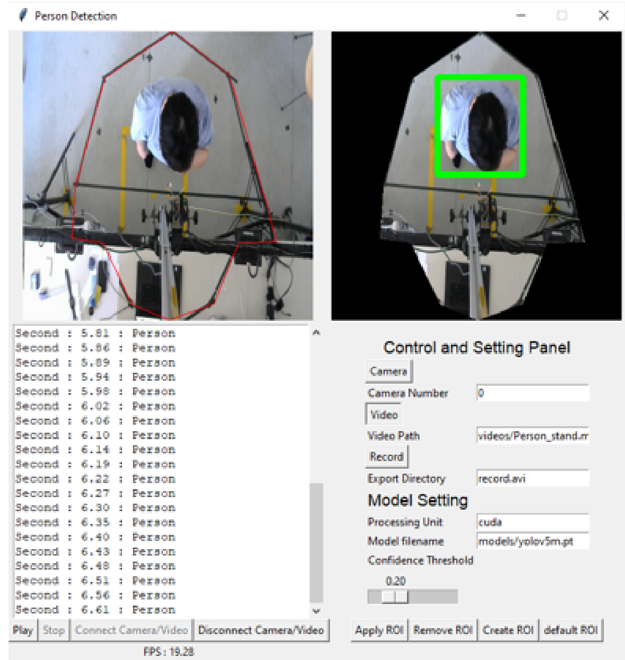

Fig. 7. Person detection software

supports GPU computing, if one is available on the machine. User can control program behaviour by changing configuration file "config.json". User can create ROI in the software and apply to the frame processing. The information on how to use software can be found in [8].

2) Export label format: When using the record mode, there will be prediction label and timestamps generated. The exported label format is as mm/dd/yy, hh:mm:ss : prediction result format. An example of labels is illustrated as in Fig. 8.

```
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:51 : Person
07/27/2022, 16:14:52 : Person
07/27/2022, 16:14:52 : Person
07/27/2022, 16:14:52 : Person
07/27/2022, 16:14:52 : Person
07/27/2022, 16:14:52 : Person
07/27/2022, 16:14:52 : Person
07/27/2022, 16:14:52 : Person
```

Fig. 8. An example of label from the software

C. Ultrasonic sensor and Red Pitaya

The Red Pitaya device model is STEMlab 125-14 and the lltrasonic sensor specification is SRF02 from Devantech.

Fig. 9 illustrates the Ultrasonic sensor mounted on Red Pitaya device.

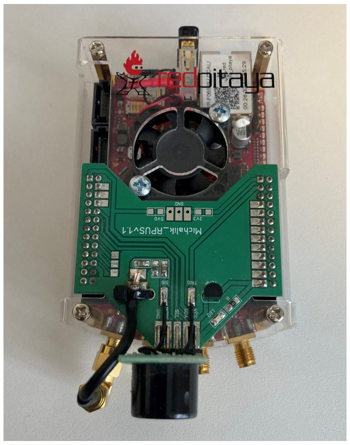

Fig. 9. Ultrasonic sensor SRF02 and Red Pitaya STEM125-14

D. Ultrasonic Data Collection Software

The Ultrasonic Data Collection Software was developed by students(s) in the Master of Information Technology Engineering. (Fig. 10) illustrates the GUI of the software. The software was set to measure analogue data or raw data from the ultrasonic sensor.

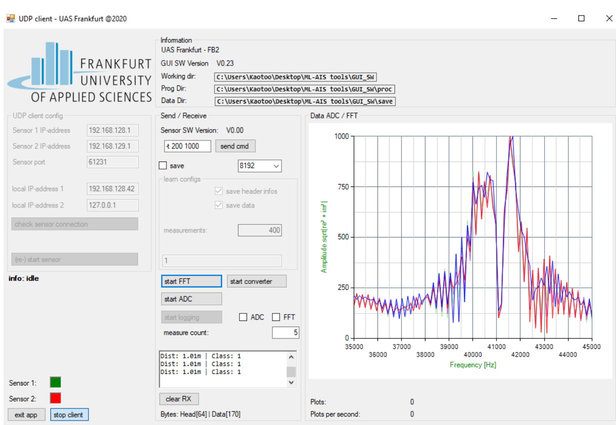

Fig. 10. The Ultrasonic Data Collection Software

E. Ultrasonic Data Format

Each data holds rows of amplitude value, ranging from 0 to 1000 and columns of frequency value, ranging from 34.9 kHz to 44.9 kHz with intervals of 1.08 1.09 kHz, which results in a total of 85 columns.

Apart from FFT data, there are also data headers exported from the software as well. Data headers contain useful information, for example, the data length, classification result from existed model in the software, or sampling frequency. The format of the data is as illustrated in Fig.11 and shown in Fig. 12.

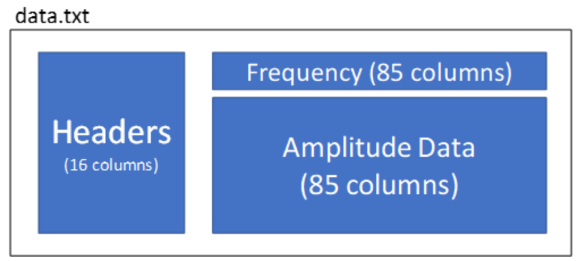

Fig. 11. Data format

34928	35048	35167	35286	35405	35524	35644	35763	35882	36001	36120	36240	36359	3647										
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	40	84	78	40	90	58	55	94
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	40	84	78	40	90	58	55	94
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	52	71	90	41	82	78	36	94
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	44	73	89	35	85	73	38	95
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	36	78	78	34	87	59	46	92
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	37	82	75	38	90	55	54	93
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	41	74	85	34	87	70	41	94
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	40	90	67	49	91	44	70	90
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	43	76	84	37	88	69	42	95
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	38	87	75	44	94	54	60	95
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	38	84	77	40	92	56	57	94
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	50	67	95	35	85	85	29	97
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	49	69	94	36	86	82	32	98
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	55	62	98	40	82	90	28	95
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	50	72	96	37	88	85	34	98
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	41	82	88	35	93	69	47	101
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	42	78	92	34	92	75	40	100
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	99	42	86	95	35	105	82	52
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	52	70	98	39	86	87	32	98
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	58	60	98	44	79	93	28	94
64	170	1	2	119	0	1953125	12	0.0	0	0.79	0	0	V0.2	0	0	59	58	96	43	76	93	27	93

Fig. 12. Raw data

In each data file, the first 16 columns of data are headers, which recorded as the setting of the measurement.Its explanation can be found as the following Table. I

TABLE I
HEADERS DESCRIPTION

Column	Header Description	Value	Remarks
1	Header lengths	64	
2	Data length	170	
3	Class detected	1 or 2	1: Object 2: Human
4	Measurement type	0 or 1	0 : FFT 1 : ADC
5	Frequency resolution f or sampling time t depend on measurement type	119	
6	-	0	irrelevant
7	Sampling frequency (Hz)	1953125	
8	ADC resolution	12	
9		0.0	irrelevant
10	Distance between sensor and first object (round-trip-time in µs)	-	
11	FFT Window length	0	
12		0	irrelevant
13	software version (RP)	V0.2	
14	aux 1	0	irrelevant
15	aux 2	0	irrelevant
16	-	-	irrelevant

III. METHODOLOGY

This section describes the methodology of the experiment. Fig. 13 illustrates the pipeline of process. The process begins with data and label collection from two sub systems. Then synchronisation between data and labels will be done using timestamps from both data. The data with label then can be transferred to the optimisation process.

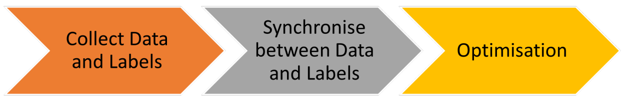

Fig. 13. Process pipeline

IV. Evaluation and Discussion

The person detection software is evaluated with the following measurements derived from [9], [10], which uses confusion matrix to evaluate predictions and actual data. Confusion matrix consists of four values. True Positive (TP) represents the value of correct predictions of positive out of actual positive cases. False Positive (FP) represents the value of incorrect positive predictions. True Negative (TN) represents the value of correct predictions of negatives out of actual negative cases. False Negative (FN) represents the value of incorrect negative predictions.

1) Accuracy: Accuracy indicates the CNN model's performance on prediction. It is calculated as the ratio between correct prediction and overall prediction. The accuracy is measured by the equation from (2).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

2) Precision: Precision indicates the CNN model's ability in order to predict positive predictions correctly out of all positive predictions. Precision is measured by the equation (3). Increasing precision means increasing correctness of the positive prediction.

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

3) Recall: Recall is similar to precision, however recall measures the correctly positive predictions out of correct prediction. Recall is measured by the equation (4). Increasing recall value means minimising the incorrect predictions

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

4) F1-Score: F1-Score involves precision and recall. It balances between both measurements. F1-Score is measured by the equation (5). F1-Score is useful in the scenario where model gives high FP and FN.

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{5}$$

In this experiment, the true positive class is person, while the true negative class is empty-scene. Three videos of different situation in the same scene were used to evaluate the software performance, which are a single person stand and move in and out , a single person sit on a chair and move, and a single chair move. These videos for evaluation and ground truth can be found in [8] under the evaluation folder. The evaluation results then were combined together as a single confusion matrix as illustrated in Fig. 14

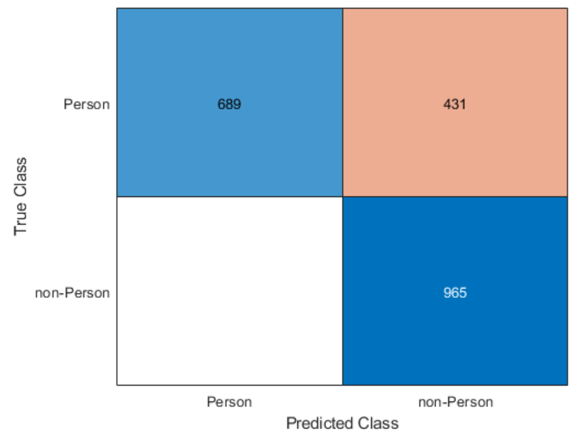

Fig. 14. Confusion matrix between person/non-person situations

Table II shows the overall evaluation scores.

TABLE II
Evaluation score

Measure	Evaluation score
Accuracy	0.79
Precision	1.00
Recall (TPR)	0.62
F1-Score	0.76
Specificity	1.00
Negative Predictive Value	0.69
False Positive Rate	0.00
False Discovery Rate	0.00
False Negative Rate	0.38

V. Conclusion

In summary to the project, the person detection software was able to classify person from other moving object in the region of interest. However, in order to be able to optimise the autonomous illumination sensor, the labels of data is expected to be all true and clean. In order to do so, the post process between both data and labels is one of the promising method and the time offset between devices will play a crucial role.

Even though the project is able to deliver positive outputs. It would be beneficial to synchronize the computer's clock with the Red Pitaya device in order to make improvements. As a result of such a method, the post-processing to clean data may be eliminated, reducing its complexity.

By providing the label and data for an autonomous illumination system, the project fulfills its purpose. In a subsequent step, the outputs will be fed into the CNN model of another system and validated.

REFERENCES

[1] Redmon, J., Divvala, S., Girshick, R., et al. (2016). You Only Look Once: Unified, Real-Time Object Detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27-30 June 2016.

[2] ultralytics, “ultralytics/yolov5: YOLOv5 in PyTorch ¿ ONNX ¿ CoreML ¿ TFLite,” GitHub, Jul. 31, 2022. [Online]. Available: https://github.com/ultralytics/yolov5. [Accessed: Jul. 31, 2022]

[3] Alexey Bochkovskiy, Chien-Yao Wang, and HongYuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934 [cs, eess], Apr. 2020. arXiv: 2004.10934.

[4] Yamashita, R., Nishio, M., Do, R.K.G. et al. Convolutional neural networks: an overview and application in radiology. Insights Imaging 9, 611–629 (2018). https://doi.org/10.1007/s13244-018-0639-9

[5] Fukushima, K. Cognitron: A self-organizing multilayered neural network. Biol. Cybernetics 20, 121–136 (1975). https://doi.org/10.1007/BF00342633

[6] A. H. Pech, P. M. Nauth and R. Michalik, ”A new Approach for Pedestrian Detection in Vehicles by Ultrasonic Signal Analysis,” IEEE EUROCON 2019 -18th International Conference on Smart Technologies, 2019, pp. 1-5, doi: 10.1109/EUROCON.2019.8861933.

[7] P. M. Nauth, A. H. Pech and R. Michalik, ”Research on a new Smart Pedestrian Detection Sensor for Vehicles,” 2019 IEEE Sensors Applications Symposium (SAS), 2019, pp. 1-5, doi: 10.1109/SAS.2019.8705978.

[8] paweenp, “paweenp/Using-Image-Processing-for-Autonomous-Illumination-Sensor-Optimization,” GitHub, 2022. https://github.com/paweenp/Using-Image-Processing-for-Autonomous-Illumination-Sensor-Optimization (accessed Jul. 31, 2022).

[9] Ahmed Fawzy Gad, ”Evaluating Deep Learning Models: The Confusion Matrix, Accuracy, Precision, and Recall”, published on 2020, accessed on 25 Aug 2021.

[10] Ajitesh Kumar, ”Accuracy, Precision, Recall & F1-Score”, published on 3 September 2021, accessed on 20 September 2021.