Optimization Method for Automated Pedestrian Detection in Autonomous Driving Based on Machine Learning

Qing Rong TANG, Xiao Fang WANG, Yuan LIAO*
Geely University of China, Chengjian Avenue No.123, Section 2, Cheng Du, 641423, China
tangqingrong@guc.edu.cn, wangxiaofang@guc.edu.cn,
luaoyuan@guc.edu.cn (**Corresponding author*)

Abstract: The rapid development of the autonomous driving technology relies on the breakthroughs in several core technologies, pedestrian detection being a critical component which directly affects the safety and reliability of autonomous driving systems. Targeting the shortcomings inherent to the current pedestrian detection technologies, this paper proposes an optimization algorithm based on machine learning, which is intended to improve the pedestrian detection accuracy and real-time efficiency. In order to optimize the parameter configuration of deep learning models, this study validates the proposed approach on two public datasets. The obtained results demonstrate that the optimized model achieved significant improvements in both pedestrian detection accuracy and computational efficiency.

Keywords: Autonomous Driving, Pedestrian Detection, Machine Learning, Optimization Algorithm.

1. Introduction

Autonomous driving technology, as a critical component of future intelligent transportation systems, has become a global focus in the fields of technology and the automotive industry. Its development trajectory spans from simple assisted driving to fully autonomous driving. The pedestrian recognition technology plays a crucial role in autonomous driving systems, as it directly impacts a vehicle's ability to perceive and make decisions in complex road environments, as well as the pedestrian safety and traffic order maintenance.

Up to this moment, the pedestrian recognition technology has achieved a significant progress. Traditional methods such as HOG (Histogram of Oriented Gradients) combined with SVM (Support Vector Machine) have achieved some success in certain application scenarios. However, they are still face challenges such as a low recognition accuracy and a poor adaptability in complex environments. Deep learning-based methods, such as YOLO (You Only Look Once) and Faster R-CNN (Region Convolutional Neural Network), have significantly improved the pedestrian recognition accuracy but came with high computational costs and a limited real-time performance (He et al., 2016; Ren et al., 2015).

Despite the above-mentioned advancements, the existing methods have obvious shortcomings. The traditional methods often perform poorly in variable environments. Deep learning methods, while accurate, have a high computational complexity, which makes them difficult to apply in scenarios with high real-time requirements. Therefore, improving the real-time performance while maintaining the recognition accuracy has become a key research topic.

This paper proposes an optimization strategy based on genetic algorithms for enhancing the accuracy and efficiency of pedestrian recognition by optimizing the parameters of deep learning models. Genetic algorithms, with their strong global search capabilities and robustness, can improve a model`s performance without significantly increasing computational costs. The significance of this research lies in providing a more reliable pedestrian recognition solution for the development of autonomous driving technology, thereby enhancing the safety and practicality of autonomous driving.

The remainder of this paper is structured as follows. Section 2 delineates the research methodology, detailing the integration of genetic algorithms with Faster R-CNN for parameter optimization. Section 3 specifies the experimental setup, including data preprocessing procedures for the Caltech and Cityscapes datasets and the baseline model configuration. Further on, Section 4 presents the empirical results obtained based on quantitative metrics (precision, recall, the F1-score, processing time) and comparative visualizations and discusses the performance improvements and computational efficiency gains, supported by a statistical analysis. Finally, Section 5 concludes this paper and proposes adaptive parameter tuning and multi-task learning as future directions.

https://doi.org/10.24846/v34i2y202508

2. Research Methods

Significant research has been conducted by scholars worldwide on pedestrian recognition in autonomous driving. Deep learning methods, represented by YOLO and Faster R-CNN, exhibit a superior performance but they are facing challenges with regard to real-time performance and computational costs. Some studies have attempted to reduce the computational load through techniques such as model pruning and quantization, but the results are limited. This problem was approached by optimizing deep learning model parameters using a genetic algorithm (GA), with the aim to improve both the recognition accuracy and efficiency (Girshick, 2015).

2.1 Overview of Machine Learning

Machine learning, a vital branch of artificial intelligence, involves building mathematical models that enable computers to learn patterns from data and make predictions. Machine learning algorithms are categorized into supervised learning, unsupervised learning, and reinforcement learning algorithms, supervised learning algorithms being the most widely used in image recognition. Deep learning, a subset of machine learning, simulates the human brain cognitive function through multi-layer neural networks and has made groundbreaking progress in fields such as image and speech recognition.

2.2 Pedestrian Recognition Technology

The evolution of the pedestrian recognition technology has progressed from traditional methods to deep learning methods. The traditional methods, such as HOG combined with SVM, achieved notable results early on but they involved complex feature extraction processes and featured a poor robustness. Deep learning methods like YOLO and Faster R-CNN, as they utilize end-to-end learning models, significantly improved the pedestrian recognition accuracy. However, these methods typically require substantial computational resources, which makes them challenging to apply in resource-constrained embedded systems (Ren et al., 2015).

2.3 Optimization Algorithms

Optimization algorithms play a critical role in tuning machine learning model parameters. The genetic algorithm (GA), a global optimization algorithm based on natural selection and genetic mechanisms, features strong global search capabilities and a high robustness. GA optimizes the objective function through selection, crossover, and mutation operations, which makes it an effective method for solving complex optimization problems.

2.4 Existing Methods

The current pedestrian recognition methods mainly include traditional methods and deep learning-based methods.

The traditional Methods mainly include HOG (Histogram of Oriented Gradients) in combination with SVM (Support Vector Machine). These methods rely on manually designed features for classification but perform poorly in complex environments, with a low recognition accuracy and adaptability.

Deep Learning Methods, such as YOLO (You Only Look Once) and Faster R-CNN (Region Convolutional Neural Network) use end-to-end learning models to automatically extract features, thereby significantly improving the pedestrian recognition accuracy. However, these methods typically require substantial computational resources and have a high computational complexity, which make them challenging to apply in resource-constrained embedded systems.

2.5 Proposed Method

This paper proposes a method based on genetic algorithm optimization of deep learning models, selecting YOLO or Faster R-CNN as the baseline model. The model parameters, such as the learning rate, number of layers, and number of nodes, are encoded into chromosomes in the genetic algorithm. The recognition accuracy and computational efficiency on the validation set are used as the fitness function. Through selection, crossover, and mutation operations, new generations of model parameters are created, and the model is trained with the new parameters and evaluated on the validation set. This process is repeated until the fitness function converges or the predetermined number of iterations is reached.

This research builds on the foundation of several important works. The YOLO method proposed by Redmon et al (2016) and the Faster R-CNN method proposed by Girshick (2015) provide the basis for applying deep learning for pedestrian recognition. Additionally, the particle swarm optimization method by Kennedy & Eberhart (1995) and the neuroevolution methods by Stanley & Miikkulainen (2002) offer theoretical support for the proposed genetic algorithm optimization (Li et al, 2022).

3. Data Collection and Processing

3.1 Datasets and Preprocessing

This study uses the Caltech Pedestrian Dataset and the Cityscapes Dataset as the primary datasets. The Caltech Pedestrian Dataset contains pedestrian images in various scenarios, suitable for pedestrian detection and recognition tasks. The Cityscapes Dataset provides rich urban street scene images, being useful for testing a model`s performance in complex environments (Wang et al., 2020). Data preprocessing steps include:

- Image Normalization: Image pixel values are normalized to the range [0, 1];

- Data Augmentation: More samples are generated through random cropping, rotation, flipping, and other methods to enhance the model's generalization ability.

3.2 Baseline Model

This paper employed Faster R-CNN as the base pedestrian recognition model. Faster R-CNN generates candidate regions through a Region Proposal Network (RPN) and then uses a Convolutional Neural Network (CNN) for feature extraction and classification. Its structure includes a backbone network (such as ResNet), a RPN, and RoI Pooling layers. The specific structure of this model is as follows:

- Backbone Network: A pre-trained ResNet-50 network is used as the feature extractor;

- Region Proposal Network (RPN): Candidate regions are generated. Given an input image I, RPN generates a set of region proposals {(xi, yi, wi, hi)};

- ROI Pooling (Region of interest): Each candidate region is mapped to a fixed-size feature map, followed by fully connected layers for classification and bounding box regression.

3.3 Optimization Algorithm Design

To optimize the parameters of the Faster R-CNN model, this paper proposes an optimization strategy based on genetic algorithms (GAs). GA optimizes the objective function through encoding, selection, crossover, and mutation operations. The specific implementation steps are as follows:

- Encoding: Model parameters (such as learning rate, weight initialization parameters etc.) are encoded into chromosomes in the genetic algorithm. Supposing there are n parameters, the chromosome is represented as $C = (p_1, p_2, ..., p_n)$;

- Fitness Function: The recognition accuracy of the model on the validation set is defined as the fitness function. The fitness function f is represented as:

$$f(C) = Accuracy(C) \tag{1}$$

where C is a parameter combination.

Selection: Excellent individuals are selected for reproduction based on fitness values. Roulette selection or tournament selection are used.

Crossover: New individuals are generated through the crossover operation. It is assumed that there are two parent individuals C_1 and C_2, and the crossover operation generates the offspring individuals $C_1^{'}$ and $C_2^{'}$:

$$C_1^{'} = (p_{1,1}, p_{2,2}, ..., p_{k,2}, p_{k+1,1}, ...p_{n,1}) \tag{2}$$

$$C_2^{'} = (p_{1,2}, p_{2,1}, ..., p_{k,2}, p_{k+1,2}, ...p_{n,2}) \tag{3}$$

where k is the intersection point.

Randomness is introduced through mutation operations to prevent local optimality. It is assumed that the mutation probability is Pm, and a gene in the chromosome is randomly selected for mutation:

$$p_i^{'} = p_i + \delta \tag{4}$$

where δ is a small-scale random disturbance.

The selection, crossover and mutation operations should be repeated until the termination condition is met (namely, the maximum number of iterations is reached or the fitness value is no longer significantly improved).

3.4 Integrated Optimization Strategy

The parameter tuning process for applying the genetic algorithm to the Faster R-CNN model is shown in Figure 1.

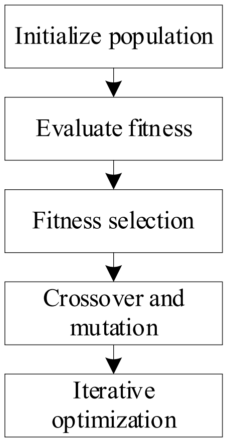

Figure 1. Parameter tuning process

The specific optimization steps are as follows:

- Initialize the population: randomly generate a set of initial parameter combinations as the initial individuals of the population;
- Evaluate fitness: calculate the recognition accuracy for each individual on the validation set as the fitness value;
- Selection operation: select excellent individuals for reproduction according to the fitness value;
- Crossover operation: perform the crossover operation on offspring individuals to generate new parameter combinations;
- Mutation operation: perform the mutation operation on some individuals to introduce random perturbations;
- Iterative optimization: repeat the above steps until the termination condition is met.

Through the above steps, the best performance parameter combination is finally obtained, and the optimized Faster R-CNN model has significantly improved with regard to pedestrian recognition tasks. This optimization algorithm can effectively tune the parameters of the deep learning model through the global search capability of the genetic algorithm, and it can improve the accuracy and real-time performance of pedestrian recognition. In practical applications, the optimization effect can be further improved by setting the parameters of the genetic algorithm (such as population size, crossover probability, mutation probability etc.) at a reasonable level.

The fitness function can be expressed as:

$$f(C) = \frac{1}{N}\sum_{i=1}^{N}\Pi(\hat{y}_i = y_i) \tag{5}$$

In summary, by introducing the genetic algorithm for optimizing the parameters of deep learning models, this paper proved the superior performance of the optimized model in pedestrian recognition tasks on two public datasets. The next step is to verify the effectiveness of the optimization method supported by specific data and details through experiments.

The definition of the Smooth L1 Loss function is as follows:

$$L(y,\hat{y}) = \begin{cases} 0.5(y-\hat{y})^2 \\ |y-\hat{y}|-0.5 \end{cases} \tag{6}$$

wherein y represents the ground truth, $\hat{y}$ denotes the prediction and $|y-\hat{y}|$ signifies the absolute difference between the ground truth and the prediction.

When the algorithm reaches the preset maximum number of iterations, the genetic algorithm is terminated.

4. Experiments and Results Analysis

4.1 Experimental Settings

In order to verify the effectiveness of the Faster R-CNN model based on the genetic algorithm optimization proposed in this paper, experiments were conducted on the Caltech Pedestrian Dataset and Cityscapes Dataset. The experimental environment is NVIDIA Tesla V100 GPU and the experiments were carried out using the TensorFlow framework. To ensure the reliability and repeatability of the experimental results, each experiment was repeated five times and the average was taken as the final result (Liu et al., 2024).

In the context of the experiments carried out, the dataset was divided into a training set and a validation set, accounting for 80% and 20% of it, respectively. The training set was used for training the model, and the validation set was used for evaluating the performance of the model. This paper compares the performance of the baseline model (the unoptimized Faster R-CNN) with that of the optimized model (Faster R-CNN optimized by the genetic algorithm) on the two datasets (Tian et al, 2018).

4.2 Experimental Results

Data preprocessing is an important step in the experiment. Image normalization and data enhancement are used to improve the generalization ability of the model. The baseline model training adopts the standard Faster R-CNN structure for training, using randomly initialized parameters and a fixed learning rate. During the training process, the cross entropy loss function is used for classification tasks, and the smooth L1 loss function is used for bounding box regression (Tian et al., 2018).

Figure 2 illustrates the pedestrian recognition results for the test image using different algorithms. The comparative pedestrian recognition results for the improved algorithm and other one-stage object detection algorithms are shown in Figure 3.

For model evaluation, the precision, recall and F1-score are used as the main evaluation indicators. In addition, the average processing time for each frame is recorded to evaluate the real-time performance of the employed model.

The experimental results for the Caltech Pedestrian Dataset show that the optimized model outperforms the baseline model for multiple evaluation indicators. The specific data is shown in Table 1.

As it can be seen, the precision of the optimized model increased from 89.6% to 93.2%, the recall rate increased from 86.3% to 90.1%, and the F1-score increased from 87.9% to 91.6%. In terms of processing time, the average processing time for each frame of the optimized model was reduced from 0.20 seconds to 0.15 seconds, showing a good real-time performance.

The experimental results for the Cityscapes Dataset further proved the effectiveness of the optimization algorithm in complex environments. The specific data is shown in Table 2.

Figure 2. Pedestrian recognition by different algorithms before improvement

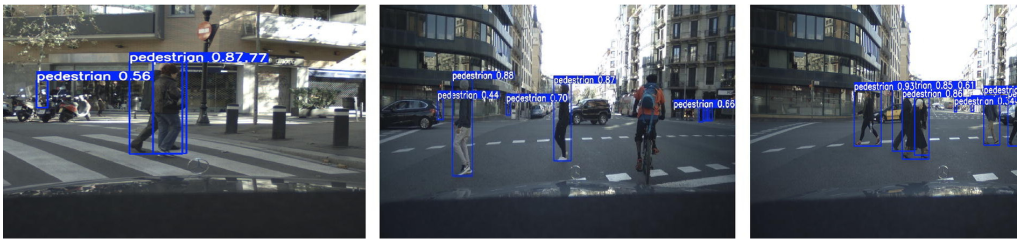

Figure 3. Pedestrian recognition by different algorithms after improvement

Table 1. Experimental results for the Caltech Pedestrian Dataset

Model	Precision	Recall	F1-score	Average processing time (seconds/frame)
Baseline model	89.6%	86.3%	87.9%	0.20
Optimized model	93.2%	90.1%	91.6%	0.15

Table 2. Experimental results for the Cityscapes Dataset

Model	Precision	Recall	F1-score	Average processing time (seconds/frame)
Baseline model	85.4%	82.1%	83.7%	0.25
Optimized model	88.9%	85.6%	87.2%	0.18

For the Cityscapes Dataset, the precision of the optimized model increased from 85.4% to 88.9%, the recall rate increased from 82.1% to 85.6%, and the F1-score increased from 83.7% to 87.2%. The average processing time was reduced from 0.25 seconds to 0.18 seconds.

4.3 Experimental Results Analysis

Through a detailed analysis of the experimental results, it can clearly be seen that the Faster R-CNN model optimized by the genetic algorithm has significantly improved the pedestrian recognition task. An in-depth discussion of the main experimental results was included below.

4.3.1 Analysis of the Optimization Effect

The experimental results for the Caltech Pedestrian Dataset and the Cityscapes Dataset show that the optimized model has achieved significant improvements for indicators such as precision, recall and F1-score. Specifically, the precision and recall rates have increased by about 3.6% and 3.8%, respectively, while the F1-score has increased by about 3.7%. These improvements are attributed to the global search capability of the genetic algorithm in the parameter space, which avoids the local optimum that traditional optimization methods are prone to reach (Tian et al., 2018).

4.3.2 Effect of Parameter Optimization

The genetic algorithm significantly improves the performance of the employed model by optimizing multiple hyperparameters of the model, such as the learning rate, weight initialization etc. Specifically, the optimized learning rate is more conducive to the rapid convergence of the model, while the adjustment of weight initialization improves the performance of the model in the early stage of training. In addition, the optimization of the data enhancement strategy enables the model to better adapt to the changes in different scenarios and improves the robustness of pedestrian recognition.

4.3.3 Analysis of Model Processing Time

The optimized model reduces the processing time for each frame by about 25%, which is particularly important with regard to the real-time requirements in practical applications. The genetic algorithm not only optimizes the accuracy and recall of the employed model, but it also significantly improves the processing speed by reducing redundant calculations and optimizing the network structure. This result shows that the optimization algorithm can effectively improve the computational efficiency of the model while also improving its performance.

The genetic algorithm has a strong global search capability and can find a solution close to the global optimal solution in a complex parameter space, which significantly improves the performance of the model. By optimizing the data enhancement strategy and model parameters, the optimized model performs more stably and features a stronger robustness in various complex scenarios. While optimizing the model performance, the genetic algorithm also reduces the computational overhead of the model and improves its processing speed, which is very important for real-time applications (Lin et al., 2020; Wang et al., 2020).

Although the optimized model is more efficient with regard to the runtime, the genetic algorithm itself requires a lot of computing resources and time during the optimization process, which may become a bottleneck in large-scale applications. The performance of the genetic algorithm is highly dependent on parameter settings, such as the population size and mutation probability. These parameters need to be tuned through a large number of experiments, which increases the complexity of the implementation. Although the genetic algorithm proved to have a significant effect in this study, its applicability and effect on other types of deep learning models need to be further verified.

5. Conclusion

This paper proposes a pedestrian recognition method based on a genetic algorithm to optimize the Faster R-CNN model. Its effectiveness on the Caltech Pedestrian Dataset and the Cityscapes Dataset is verified through experiments. The research results show that the optimized model has achieved significant improvements with regard to accuracy, the recall rate, F1-score and processing time. This shows the potential and application value of genetic algorithms in deep learning model parameter optimization (Tan et al., 2020; Aydin et al., 2023).

The main advantages of the proposed model are performance improvement, computational efficiency and an enhanced robustness. So, the optimized Faster R-CNN model shows a higher accuracy and recall rate in pedestrian recognition tasks, which significantly improves the pedestrian recognition performance. To that, while ensuring a high recognition rate, the optimized model significantly reduces the processing time and improves the real-time performance of the system. Finally, the optimized model performs more stably in complex scenarios and features a stronger adaptability and robustness.

Further research directions on genetic algorithm optimization could focus on several aspects.

First, adaptive parameter adjustment, which would involve carrying out research on adaptive genetic algorithms to enable them to dynamically adjust parameters such as the population size and mutation probability according to actual conditions, further improving the optimization efficiency.

Second, a hybrid optimization strategy could be employed, which would combine other optimization methods (such as particle swarm optimization, simulated annealing etc. (Muthu & Kalimuthu, 2023; Yehoshua et al., 2023) to form a hybrid optimization strategy, maximize the advantages of each algorithm, and improve the optimization effect.

To that, future research could focus on the implementation of larger-scale datasets and practical applications. In this context, the large-scale dataset verification and real-time system integration should be carried out. The former would consist in verifying the performance of the optimization model on a larger and more complex real-world dataset to evaluate its adaptability to different application scenarios, while the latter would focus on integrating the optimization model into the actual autonomous driving system, testing its performance in a real environment, and further optimizing its real-time performance and stability.

Finally, model structure improvement can be pursued, which would involve a lightweight model design and multi-task learning. Designing a more lightweight network structure would further reduce the computing overhead and the processing time while ensuring a high model performance and multi-task learning could involve performing the analysed tasks in combination with other related tasks (such as pedestrian attribute recognition, behavior prediction etc.) and adopting a multi-task learning method in order to improve the comprehensive ability and application value of the proposed model (Yehoshua et al., 2023; Fan et al., 2023).

It is hoped that these further studies and improvements would provide more efficient and reliable pedestrian recognition solutions for the development of the autonomous driving technology, thereby improving the overall performance and safety of autonomous driving systems.

REFERENCES

Aydin, I., Sevi, M., Akin, E. et al. (2023) A Deep Learning-Based Hybrid Approach to Detect Fastener Defects in Real-Time. *Tehnički vjesnik* [*Technical gazette*]. 30(5), 1461-1468. https://doi.org/10.17559/TV-20221020152721.

Fan, X., Li, H., Liu, X. et al. (2023) Illegal Intrusion Detection of Internet of Things Based on Deep Mining Algorithm. *Tehnički vjesnik* [*Technical gazette*]. 30(6), 1968-1974. https://doi.org/10.17559/TV-20230808000860.

Girshick, R. (2015) Fast R-CNN. In: *2015 Proceedings of the IEEE International Conference on Computer Vision, 7-13 December 2015, Santiago, Chile.* New York, USA, IEEE. pp. 1440-1448.

He, K., Zhang, X., Ren, S. et al. (2016) Deep Residual Learning for Image Recognition. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, 27-30 June 2016, Las Vegas, NV, USA.* New York, USA, IEEE. pp. 770-778.

Kennedy, J. & Eberhart, R. (1995) Particle Swarm Optimization. In: *Proceedings of ICNN'95 - International Conference on Neural Networks, 27 November - 1 December 1995, Perth, WA, Australia*. New York, USA, IEEE. pp. 1942-1948.

Li, Y., Li, M., Lu, J. et al. (2022) An Auto-processing Method of Traffic Safety Information Based on a Multi-task Transfer Learning Algorithm. *China Journal of Highway and Transport*. 35(9), 1-12. https://doi.org/10.19721/j.cnki.1001-7372.2022.09.001.

Lin, T.-Y., Goyal, P., Girshick, R. et al. (2020) Focal Loss for Dense Object Detection. *IEEE Transactions on Pattern Analysis and Machine*

Intelligence. 42(2), 318-327. https://doi.org/10.1109/TPAMI.2018.2858826.

Liu, Z., Dong, Z., Zhou, T et al. (2024) Review and Prospects of Performance Enhancement of Asphalt Mixtures Based on Material Informatics. *China Journal of Highway and Transport*. 37(4), 98-120. https://doi.org/10.19721/j.cnki.1001-7372.2024.04.009.

Redmon, J., Divvala, S., Girshick, R. et al. (2016) You Only Look Once: Unified, Real-Time Object Detection. In: *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, 27-30 June 2016, Las Vegas, NV, USA*. New York, USA, IEEE. pp. 779-788.

Ren, S., He, K., Girshick, R. et al. (2015) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: Cortes, C., Lee, D. D., Sugiyama, M. & Garnett, R. (eds.) *Advances in Neural Information Processing Systems 28 (NIPS 2015)* (*NIPS'15: Proceedings of the 29th International Conference on Neural Information Processing Systems, 7–12 December 2015, Montreal, Canada*). Cambridge, MA, USA, MIT Press, pp. 91-99.

Stanley, K. O. & Miikkulainen, R. (2002) Evolving Neural Networks through Augmenting Topologies. *Evolutionary Computation*. 10(2), 99-127.

Tan, M., Pang, R., Le, Q. V. (2020) EfficientDet: Scalable and Efficient Object Detection. In: Proceedings of the 2020 IEEE/CVF Conference on *Computer Vision and Pattern Recognition*, *13-19 June 2020, Seattle, WA, USA*. New York, USA, IEEE. pp. 10778-10787.

Tian, Y., Gelernter, J., Wang, X. et al. (2018) Lane Marking detection via deep convolutional neural network. *Neurocomputing*. 280, 46-55. https://doi.org/10.1016/j.neucom.2017.09.098

Wang, Q., Wu, B., Zhu, P. et al. (2020) ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, *13-19 June 2020, Seattle, WA, USA*. New York, USA, IEEE. pp. 11531-11539.

Yehoshua A., Bechar A., Cohen Y. et al. (2023) Dynamic Sampling Algorithm for Agriculture-Monitoring Ground Robot. *International Journal of Simulation Modelling*. 22(3), 392-403. https://doi.org/10.2507/IJSIMM22-3-646.