Implementation of Computer Vision in VIDAS for Road Damage Detection Using the SSD Algorithm

Rian Rahmanda Putra[1], Anggi Nidya Sari[2], Agum Try Wardhana[3], Yulian Mirza[4], Muhammad Fikri[6], and Norca Praditya[6]
[1,4, 5] Computer Engineering, State Polytechnic of Sriwijaya, Palembang, Indonesia
[2, 6] Civil Engineering, State Polytechnic of Sriwijaya, Palembang, Indonesia
[3] Electrical Engineering, State Polytechnic of Sriwijaya, Palembang, Indonesia
rianrahmanda@polsri.ac.id

Abstract. Detecting road damage is essential for maintaining safe and efficient transportation infrastructure. Automated systems for identifying road defects, such as cracks and potholes, play a crucial role in enhancing the speed and accuracy of road inspections, enabling timely and cost-effective maintenance. This study focuses on implementing a Single Shot Detector (SSD) model to detect various types of road damage. Built on the Mobile-Net V2 architecture with FPN-Lite, the SSD model was trained on a dataset containing images of road defects and optimized over 50,000 iterations with a batch size of 16 for real-time detection. The model's effectiveness was evaluated using the mean Average Precision (mAP) metric at a threshold of 0.5:0.95, yielding an overall mAP score of 13.22%. The results showed that the model performed best in detecting faded markings (mAP 24.77%) and alligator cracks (mAP 20.49%), while struggling with transverse cracks (mAP 1.06%) and potholes (mAP 9.18%). These findings highlight the potential of the SSD model to support real-time road maintenance applications by reliably detecting certain defect types, but also indicate areas for further enhancement. Improvements in model training and data augmentation could elevate detection accuracy across all defect categories, contributing to more effective and proactive road infrastructure monitoring. This advancement in automated road damage detection ultimately supports faster, data-driven maintenance decisions, leading to safer and longer-lasting road networks.

Keywords: Road damage detection, SSD model, MobileNet V2, road infrastructure monitoring.

1 Introduction

Road infrastructure is vital for ensuring community connectivity and mobility. The government is tasked with the responsibility of maintaining these roads [1]. Various factors contribute to road damage, such as the weight of vehicles, weather conditions, the quality of materials used, and maintenance practices. Common road defects include longitudinal cracks, transverse cracks, alligator cracking, and potholes [2]. The consequences of road damage are widespread, impacting road safety, increasing repair costs, and affecting the broader community. Disruptions to local economic activities, especially for business dependent on smooth road access for the distribution for goods and services, are significant. Additionally, road damage can

N. L. Husni et al. (eds.), *Proceedings of the 8th FIRST 2024 International Conference on Global Innovations (FIRST-ESCSI 2024)*, Advances in Engineering Research 261,
https://doi.org/10.2991/978-94-6463-678-9_55

hinder accessibility to essential services such as healthcare and education, particularly in remote areas, and contribute to environmental issues like increased greenhouse gas emissions due to worsening traffic congestion. Early detection and monitoring of road damage are therefore crucial steps in ensuring safe and sustainable road infrastructure [3].

Traditionally, road damage detection and monitoring involve surveys to gather necessary information, such as geometric, structural, and road condition data. Video-based road damage detection system offer a cost-effective and efficient alternative for road condition assessment and evaluation [4]. Several studies have explored video-based road damage detection methods, such as Road-TransTrack, which detects road cracks and potholes with 91.60% accuracy [5], EcRD for rapid and accurate road damage detection using cloud computing [6] and automated deformation detection using artificial intelligence object detection functions from video to identify cracks, joints, and repair marks [7], [8]. Despite these efforts, challenges remain, particularly in achieving high accuracy in damage detection under varying environmental conditions and the presence of closely spaced objects in the image.

Road damage can be assessed through video using computer vision techniques, notably the Single Shot Detector (SSD) algorithm. SSD is a well-known object detection algorithm commonly employed in computer vision and machine learning [9]. It functions within convolutional neural networks (CNNs) and is engineered to identify objects within an image with high precision [10]. The architecture of SSD is fully convolutional, allowing it to detect all objects in an image in a single pass, making it highly effective for real-time applications [11]. During prediction, the network assigns scores to indicate the likelihood of each object category within every bounding box, facilitating rapid and accurate object localization. Due to its speed and effectiveness, SSD has become a favored option for various computer vision tasks, including real-time detection of road damage [12].

A system for detecting road damage should be widely applicable and easily accessible to stakeholders, including local governments and road maintenance agencies. The Public Works and Spatial Planning Office (PUPR) is integral to road maintenance, performing field surveys to visually assess road conditions and identify damages like cracks and potholes. The research problem is outlined as follows: (1) How can the SSD algorithm be applied to classify road damage types using dash-cam video footage? (2) What is the accuracy of the SSD-based model in classifying different types of road damage? (3) How effective is the application in monitoring road conditions? This study is limited by the following parameters: (1) The classification model is implemented using the SSD algorithm; (2) The images used focus on road conditions, including classifications for longitudinal cracks, transverse cracks, alligator cracks, potholes, and faded markings; (3) The image dataset is generated by extracting images from labelled dash-cam video footage; (4) The dataset labelling is performed manually and validated by PUPR partners; (5) The final product is software designed to detect road damage from dash-cam video footage, with the output being a report detailing the types of road damage detected.

2 Literature Review

2.1 Computer Vision

Computer vision is a field of artificial intelligence that enables computers to mimic the visual perception abilities of humans. It allows computers to extract valuable information from digital images, videos, or other visual sources and make decisions or provide recommendations based on the extracted data [13]. In recent years, research in computer vision has advanced significantly. For instance, Liu et al. [9] introduced the Single Shot Multi-Box Detector (SSD), a real-time object detection algorithm capable of identifying multiple objects within an image with high accuracy. Later, research by Chen et al. [14] demonstrated further improvements in SSD for real-time object detection. The SSD algorithm has also been optimized for mobile devices with limited resources [15]. According to recent studies, the SSD algorithm achieves competitive accuracy levels compared to other state-of-the-art computer vision algorithms, such as RCNN, while operating at significantly higher speeds [16].

2.2 Single Shot Detector (SSD)

The Single Shot Detector (SSD) is an innovative object detection algorithm in the field of computer vision. It excels at quickly and accurately detecting and localizing objects within images or videos with a single application of a deep neural network, making SSD highly efficient and ideal for real-time applications. SSD operates by using anchor boxes with various aspect ratios at multiple positions on the feature map [9]. These anchor boxes enable effective handling of objects with diverse sizes and shapes. Additionally, SSD leverages multi-scale feature maps to detect objects across different scales. A key component adapted by SSD from the Multi-Box framework is the use of Confidence Loss and Location Loss with a smooth L1-Norm, which enhances its suitability for real-time applications, such as face detection [17] (Hu and Huang, 2020), vehicle detection [18], and pedestrian detection [19]

2.3 Mean Average Precision (mAP)

Mean Average Precision (mAP) is a metric used to evaluate the performance of object detection models in the field of computer vision [20]. This metric generally assesses a model's ability to accurately recognize and localize objects [21]. In object detection, a model is expected not only to identify the class of objects present in an image but also to pinpoint their locations using bounding boxes. As such, a metric that can effectively measure both classification accuracy and localization precision is essential, and mAP fulfills this role. mAP is calculated as the average of the Average Precision (AP) values across all object classes in a dataset, providing a comprehensive summary of the model's detection performance.

3 Research Methodology

The research methodology for the implementation of computer vision in VIDAS for road damage detection using SSD algorithm follows a systematic approach the can be seen in the figure 1.

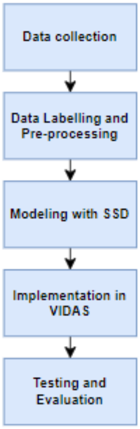

Figure 1. Research methodology for implementing computer vision in VIDAS using SSD algorithm

The research methodology begins with data collection, where video footage of road conditions is captured using dash-cams mounted on operational vehicles from the Public Works and Spatial Planning Department (PUPR). These dash-cams record various road surfaces in real-time across diverse environments, providing a broad dataset representative of actual conditions. To create a robust dataset, the recorded videos are split into individual image frames during the labelling and pre-processing phase. This phase involves manually labeling each extracted image to mark specific types of road damage, such as longitudinal cracks, transverse cracks, alligator cracks, potholes, and faded road markings. By breaking down video footage into single images, the dataset captures a wider variety of damage examples under different lighting and weather conditions, enhancing the model's adaptability.

Following labeling, the dataset undergoes pre-processing, including resizing, normalization, and data augmentation, ensuring it is ready for model training. In the modeling stage, SSD model is developed and trained to detect and classify road damage accurately. The SSD model is refined using cross-validation techniques to ensure it performs well across diverse conditions. This training phase is critical in enabling VIDAS to achieve high accuracy in detecting and classifying damage in real-time. Once the SSD model has been trained and validated, the study advances to the VIDAS integration stage, where the SSD model is embedded within the VIDAS software. This integration enables VIDAS to process video inputs directly from dash-cams, analyzing frames in real-time and visually highlighting the types and locations of detected damage.

In the final testing and evaluation phase, VIDAS's performance is evaluated mAP metric across multiple IoU thresholds. This metric helps determine the system's accuracy in various scenarios, ensuring VIDAS can reliably support infrastructure monitoring and maintenance by accurately identifying and categorizing road damage. Data collection methods are an important step in the research process to obtain the necessary information. In this study, data on disease distribution and environmental factors will be collected using an integrated approach. The following are the data collection methods that will be used:

3.1 Data Collection

The initial stage involved collecting video data of road conditions using dash-cams installed in operational vehicles from the Public Works and Spatial Planning Department (PUPR). The captured videos covered various road environments, ensuring a diverse dataset representative of typical road damage scenarios, such as cracks, potholes, and faded road markings. This comprehensive data collection was crucial for training an effective model.

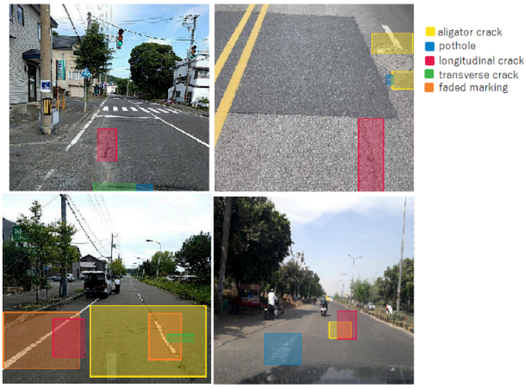

Figure 2. Image dataset for road damage detector

3.2 Data Labelling and Pre-Processing

Once the data was collected, the videos were analyzed by extracting individual frames, which were then manually labeled to identify different types of road damage.


```
<annotation>
    <folder>train</folder>
    <filename>China_Drone_000001.jpg</filename>
    <path>D:\Sayembara\dataset_dir\train\China_Drone_000001.jpg</path>
    <source>
        <database>Unknown</database>
    </source>
    <size>
        <width>512</width>
        <height>512</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name>D10</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>14</xmin>
            <ymin>70</ymin>
            <xmax>198</xmax>
            <ymax>99</ymax>
        </bndbox>
    </object>
    <object>
        <name>D00</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>89</xmin>
            <ymin>114</ymin>
            <xmax>136</xmax>
            <ymax>284</ymax>
        </bndbox>
    </object>
</annotation>
```


Figure 3. Data labelling and pre-processing

This labeling process was critical for building a strong training dataset. The labeling was performed using the LabelImg tool, with annotations saved in the Pascal VOC format. The pre-processing stage involved resizing images, normalizing pixel values, and applying data augmentation techniques to expand the diversity of the training data. These steps were designed to improve the model's ability to generalize across a range of road conditions and environments.

3.3 Modelling with SSD

The model testing was conducted to obtain an SSD model with high accuracy in classifying different types of road damage in the given video frames. The SSD model used in this study is 'SSD Mobile-Net V2 FPN-Lite, which is an SSD Mobile-Net V2 model with FPN-Lite.

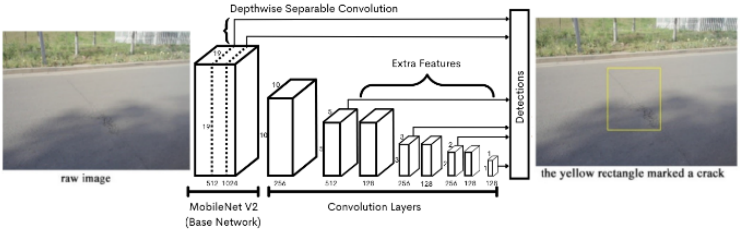

Figure 4. SSD model using MobileNet V2 for road damage detector

This model has an input image resolution of 640x640 pixels and was chosen for its suitability in object detection on devices with limited resources, given the computational constraints of the equipment available in the computer laboratory of the Computer Engineering Department. Interviews and questionnaires will be conducted with local communities to obtain information on socioeconomic and behavioral factors that could potentially affect health. This will help in further understanding of the social determinants of health in the area. The model underwent training for 50,000 iterations with a batch size of 16. This setup was chosen to speed up the training process and potentially enhance gradient stability, finding a balance between training efficiency and memory consumption, while ensuring the model could learn from adequate variation in each batch. The model was trained to identify different types of road damage, including longitudinal cracks, transverse cracks, alligator cracks, potholes, and faded road markings.

3.4 Implementation in VIDAS

The trained SSD model was then integrated into VIDAS software system. VIDAS stands for Video-Based Intelligent Damage Analysis from Sriwijaya. It was developed to provide an interface for the SSD model trained to recognize road damage, making it user-friendly for end users. The architecture of VIDAS, as shown in the figure 5, involves several key components working together to analyze road damage from video input.

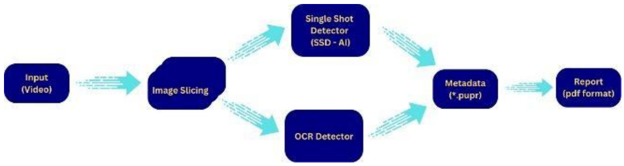

Figure 5. SSD model using MobileNet V2 for road damage detector

Based on the figure 5, the key component of VIDAS are:

1. Input (Video): The process begins with a video input, typically captured from a dash-cam or similar device during road inspections.
2. Image Slicing: The video is then sliced into individual frames or images. These images are processed separately to extract relevant information.
3. Single Shot Detector (SSD-model): The segmented images are fed into the Single Shot Detector (SSD), an AI-based model designed to recognize different kinds of road damage, including cracks and potholes. The SSD model identify these images to detect and categorize the damage.
4. OCR detector: parallel to the SSD, an optical character recognition (OCR) detector is employed to extracted any textual information from the images, such as road signs or markings the might be relevant to the damage report.
5. Metadata generation: The results from the SSD and OCR detector are combined into metadata, providing detailed information about the type and location of road damage.

The figure 6 shows user interface (UI) of the VIDAS software, specifically the section for capturing and analysing road damage from video files.

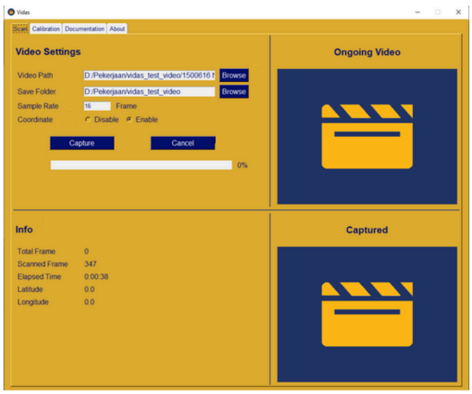

Figure 6. UI of VIDAS Software

The road damage report generated by the VIDAS application is illustrated in figure 7.

Figure 7. The road damage report generated by the VIDAS application

4 Result and Discussion

The performance evaluation of our proposed SSD-based road damage detection model demonstrates promising results across various damage categories. As shown in Table 1, the model achieved an overall mAP value of 13.22% at IoU thresholds of 0.5 and 0.85, indicating its capability to detect multiple types of road damages under strict evaluation criteria.

Referring to Table 1, the model exhibited particularly strong performance in detecting faded markings and alligator cracks, with mAP scores of 24.77% and 20.49% respectively. This impressive detection rate for these critical damage types is especially valuable since alligator cracks often indicate structural deterioration of the pavement, while faded markings pose significant safety concerns for road users. The model's proficiency in identifying these defects could potentially enable more timely maintenance interventions.

Table 1. The performance results of the SSD model

Class	Average mAP @ 0.5 : 0.95
Longitudinal cracks	10.60%
Transverse cracks	1.06%
Alligator cracks	20.49%
Potholes	9.18%
Faded markings	24.77%
Overall	**13.22%**

Based on Table 1, for longitudinal cracks, the model achieved a respectable mAP of 10.60%, while pothole detection showed a reasonable performance with 9.18% mAP. Given that these damage types often have varying appearances and sizes, these results suggest that our model has developed a robust capability to recognize diverse damage patterns. The relatively lower performance on transverse cracks (1.06%) presents an opportunity for future improvements, possibly through data augmentation or architectural modifications. The varying performance across different damage categories provides valuable insights for future model enhancements. The strong performance in detecting complex patterns like alligator cracks suggests that the model has successfully learned to identify intricate damage patterns. This capability could be leveraged to improve detection rates for other damage types through targeted training strategies. These results are particularly encouraging considering the challenging nature of road damage detection, which often involves dealing with varying lighting conditions, weather effects, and complex damage patterns. The model's ability to maintain consistent detection capabilities across multiple damage categories demonstrates its potential for practical implementation in automated road maintenance systems.

5 Conclusion

This study presents a comprehensive evaluation of an SSD model's capabilities in road defect detection. The experimental results, as detailed in Table 1, reveal promising outcomes across different defect classes, with an overall mAP of 13.22%. The model demonstrates particular strength in detecting complex patterns such as alligator cracks (20.49% mAP) and faded markings (24.77% mAP), showcasing its potential for automated road condition assessment. While the detection of longitudinal cracks (10.60% mAP) shows encouraging results, the current performance on transverse cracks (1.06% mAP) indicates an opportunity for targeted improvements. This variation in detection accuracy across different defect types provides valuable insights for optimizing the model's real-world application in road infrastructure monitoring.

To further enhance the model’s practical utility, several promising research directions have been identified:

1. Strategic expansion of the training dataset, particularly focusing on challenging cases like transverse cracks and potholes, to strengthen the model's ability to handle diverse real-world scenarios.
2. Integration of complementary detection algorithms that could work in conjunction with the SSD model, potentially creating a more robust system capable of addressing the current limitations in detecting specific defect types.
3. Advanced optimization of the model's architecture and parameters, specifically targeting improved detection accuracy for defect classes currently showing lower mAP scores.

These enhancement strategies hold significant potential for elevating the SSD model's performance in practical applications. By addressing the current detection challenges, particularly in classes with lower mAP scores, the model can evolve into a more comprehensive tool for road condition monitoring. The implementation of these improvements could lead to a more reliable and efficient system for infrastructure maintenance, ultimately contributing to enhanced road safety and optimized resource allocation in road maintenance operations.

References

[1] S. E. Priana, ‘ANALISA FAKTOR PENYEBAB KERUSAKAN JALAN (STUDI KASUS RUAS JALAN LINGKAR UTARA KOTA PADANG PANJANG)’, *RTJ*, vol. 1, no. 1, Apr. 2018, doi: 10.31869/rtj.v1i1.609.

[2] A. W. O. Gama, D. A. P. A. G. Putri, and G. H. Prathama, ‘Sistem Pakar Untuk Mendeteksi Jenis Kerusakan Jalan: Studi Kasus pada Perkerasan Lentur’, *tc*, vol. 21, no. 3, pp. 554–564, Aug. 2022, doi: 10.33633/tc.v21i3.6145.

[3] J. O. Simanjuntak, D. Suita, M. H. M. Hasibuan, N. I. M. Simanjuntak, and W. R. Halawa, ‘KERUSAKAN JALAN YANG BERIMBAS PADA LINGKUNGAN DAN PENGGUNA (PEMAKAI) JALAN RAYA DILOKASI JALAN VETERAN, HELVETIA, MEDAN’, *JUDA*, vol. 31, no. 3, p. 11, Jun. 2023, doi: 10.46930/ojsuda.v31i3.3354.

[4] F. Kortmann *et al.*, ‘Detecting Various Road Damage Types in Global Countries Utilizing Faster R-CNN’, in *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA: IEEE, Dec. 2020, pp. 5563–5571. doi: 10.1109/BigData50022.2020.9378245.

[5] N. Wang, L. Shang, and X. Song, ‘A Transformer-Optimized Deep Learning Network for Road Damage Detection and Tracking’, *Sensors*, vol. 23, no. 17, p. 7395, Aug. 2023, doi: 10.3390/s23177395.

[6] Y. Yuan, Md. S. Islam, Y. Yuan, S. Wang, T. Baker, and L. M. Kolbe, ‘EcRD: Edge-Cloud Computing Framework for Smart Road Damage Detection and Warning’, *IEEE Internet Things J.*, vol. 8, no. 16, pp. 12734–12747, Aug. 2021, doi: 10.1109/JIOT.2020.3024885.

[7] H. Emoto, M. Numata, and A. Shiga, ‘Application of Ai-based Deformation Extract Function from a Road Surface Video to a Road Pavement Condition Assessment System’, in *Proceedings of The 17th East Asian-Pacific Conference on Structural Engineering and Construction, 2022*, vol. 302, G. Geng, X. Qian, L. H. Poh, and S. D. Pang, Eds., in Lecture Notes in Civil Engineering, vol. 302. , Singapore: Springer Nature Singapore, 2023, pp. 1503–1514. doi: 10.1007/978-981-19-7331-4_119.

[8] K. Doshi and Y. Yilmaz, ‘Road Damage Detection using Deep Ensemble Learning’, in *2020 IEEE International Conference on Big Data (Big Data)*, Atlanta, GA, USA: IEEE, Dec. 2020, pp. 5540–5544. doi: 10.1109/BigData50022.2020.9377774.

[9] W. Liu *et al.*, ‘SSD: Single Shot MultiBox Detector’, 2015, doi: 10.48550/ARXIV.1512.02325.

[10] J. Deng, Y. Pan, T. Yao, W. Zhou, H. Li, and T. Mei, ‘Single Shot Video Object Detector’, *IEEE Trans. Multimedia*, vol. 23, pp. 846–858, 2021, doi: 10.1109/TMM.2020.2990070.

[11] A. Kumar, Z. J. Zhang, and H. Lyu, ‘Object detection in real time based on improved single shot multi-box detector algorithm’, *J Wireless Com Network*, vol. 2020, no. 1, p. 204, Dec. 2020, doi: 10.1186/s13638-020-01826-x.

[12] X. Wang *et al.*, ‘Single shot multibox detector object detection based on attention mechanism and feature fusion’, *J. Electron. Imag.*, vol. 32, no. 02, Apr. 2023, doi: 10.1117/1.JEI.32.2.023032.

[13] A. Khan, A. Laghari, and S. Awan, 'Machine Learning in Computer Vision: A Review', *ICST Transactions on Scalable Information Systems*, p. 169418, Jul. 2018, doi: 10.4108/eai.21-4-2021.169418.
[14] W. Chen, Y. Qiao, and Y. Li, 'Inception-SSD: An improved single shot detector for vehicle detection', *J Ambient Intell Human Comput*, vol. 13, no. 11, pp. 5047–5053, Nov. 2022, doi: 10.1007/s12652-020-02085-w.
[15] J. Zhang, J. Jing, P. Lu, and S. Song, 'Improved MobileNetV2-SSDLite for automatic fabric defect detection system based on cloud-edge computing', *Measurement*, vol. 201, p. 111665, Sep. 2022, doi: 10.1016/j.measurement.2022.111665.
[16] Y. Liu, R. Liu, S. Wang, D. Yan, B. Peng, and T. Zhang, 'Video Face Detection Based on Improved SSD Model and Target Tracking Algorithm', *JWE*, Jan. 2022, doi: 10.13052/jwe1540-9589.21218.
[17] X. Hu and B. Huang, 'Face Detection based on SSD and CamShift', in *2020 IEEE 9th Joint International Information Technology and Artificial Intelligence Conference (ITAIC)*, Chongqing, China: IEEE, Dec. 2020, pp. 2324–2328. doi: 10.1109/ITAIC49862.2020.9339094.
[18] J. Cao *et al.*, 'Front Vehicle Detection Algorithm for Smart Car Based on Improved SSD Model', *Sensors*, vol. 20, no. 16, p. 4646, Aug. 2020, doi: 10.3390/s20164646.
[19] D. Liu, S. Gao, W. Chi, and D. Fan, 'Pedestrian detection algorithm based on improved SSD', *IJCAT*, vol. 65, no. 1, p. 25, 2021, doi: 10.1504/IJCAT.2021.113643.
[20] B. Wang, 'A Parallel Implementation of Computing Mean Average Precision', 2022, *arXiv*. doi: 10.48550/ARXIV.2206.09504.
[21] S. Prasanna and M. El-Sharkawy, 'Improving Mean Average Precision (mAP) of Camera and Radar Fusion Network for Object Detection Using Radar Augmentation', in *Proceedings of Seventh International Congress on Information and Communication Technology*, vol. 465, X.-S. Yang, S. Sherratt, N. Dey, and A. Joshi, Eds., in Lecture Notes in Networks and Systems, vol. 465. , Singapore: Springer Nature Singapore, 2023, pp. 51–60. doi: 10.1007/978-981-19-2397-5_6.