

Road Damage Detection Using Yolov9-Based Imagery

Febrian Akbar Azhari ^[1], Tatang Rohana ^[2], Kiki Ahmad Baihaqi ^{*[3]}, Ahmad Fauzi ^[4]

Department of Computer Science ^{[1], [2], [3], [4]}

University of Buana Perjuangan Karawang
Karawang, Indonesia

if21.febrianazhari@mhs.ubpkarawang.ac.id^[1], tatang.rohana@ubpkarawang.ac.id^[2], kikiyahmad@ubpkarawang.ac.id^[3],
afauzi@ubpkarawang.ac.id^[4]

Abstract— Road damage is one of the leading factors contributing to traffic accidents. Rapid identification and repair of damaged roads are crucial in road infrastructure management. This study aims to develop an effective method for detecting road damage, utilizing the YOLOv9 algorithm as a key component, such as cracks and potholes, using the Convolutional Neural Network (CNN) approach. YOLOv9 was chosen due to its efficient architecture, which enables real-time object detection, and its proven effectiveness in various object detection tasks. An annotated dataset of road images was used during the model training and testing process. The results show that the YOLOv9 model can accurately detect road damage. The model achieved a precision of 0.85 and a recall of 0.992 for pothole detection, and a precision of 0.94 for crack detection. Evaluation using mAP50 yielded a score of 0.96, while mAP50-95 reached 0.77, indicating strong detection and classification capability. A consistent decline in loss functions during training also signifies effective learning by the model. These findings suggest that YOLOv9 has the potential to be implemented in automated road damage detection systems, which can accelerate maintenance processes and enhance road user safety.

Keywords: YOLOv9, Road Damage Detection, CNN, Deep Learning

I. INTRODUCTION

Highways have a vital role as infrastructure to support strategic development [1], which requires effective management. Efforts to reduce potential hazards in driving on the highway are very important [2]. Various factors can cause road accidents, one of which is the condition of damaged roads [3]. Proper maintenance of the road network is one of the important steps in reducing the risk of danger to motorists. Given that roads are a major part of the crucial land transportation infrastructure [4], [5], the maintenance of the road network needs to be managed in a sustainable manner [6].

Monitoring road conditions is a very important aspect to minimize the number of accidents caused by bad roads. According to [3], accidents occur due to various factors, and the most dominant factor is undetected or unrepaired road defects. Therefore, monitoring of a large road network is essential to determine the appropriate and different types of maintenance for each section of the road [7].

Continuous monitoring in a region with many roadways presents a challenge, as field surveys must be conducted at various locations throughout the road network. Manual methods used to identify road defects are time-consuming. However, with the development of computer technology, many jobs that previously required human labor can now be replaced by artificial intelligence-based systems.

The development of artificial intelligence has a major role in the future transformation of various industries [8]. In this research, the author adopts the Convolutional Neural Networks (CNN) algorithm model for digital image processing. The CNN used in this study proved to be effective in identifying objects present in digital images of the ground surface [9]. By using cameras and applying the CNN model, the process of field surveys to detect road damage can be carried out more quickly, which allows repair of damaged roads to be carried out more immediately, thereby reducing the number of accidents caused by poor road conditions.

The study conducted by [10] investigated the use of camera image processing technology to automatically detect road defects. The study by [10] applied an artificial neural network-based method using the Mask R-CNN algorithm was applied to solve the road defect problem, perform classification, as well as extract important features in the image. The research conducted by [11] used pavement image data taken using an Unmanned Aerial Vehicle (UAV). They managed to distinguish between normal and damaged pavements, including detecting cracks and potholes. The results of this study show that remote sensing systems using UAVs can provide an effective tool for monitoring the condition of asphalt pavements.

Previous studies have demonstrated the success of Convolutional Neural Networks (CNNs) in various image classification and object detection tasks. [12] utilized CNNs to distinguish between damaged (with holes, tears, and dents) and undamaged cardboard packaging, achieving a high accuracy of 95.77%. [13] also employed CNNs to classify five types of mangoes, resulting in a 99.56% accuracy. Furthermore, [14] applied CNNs to recognize Sundanese script patterns, with varying accuracy depending on the image source, reaching 100% for computer font images. The You Only Look Once (YOLO) method, frequently used to test CNN architectures, has also been successfully implemented in detecting rat pests in

farmland using YOLOv5 (88% accuracy) [15] and classifying rice types using YOLOv3, with results differing based on rice stacking (60% when stacked, 100% when not) [16]. These studies highlight the broad applicability of CNNs and YOLO-based methods in image analysis and object detection, providing a foundation and motivation for employing YOLOv9 in road damage detection.

Although several studies have applied CNN-based methods such as Mask R-CNN or earlier versions of YOLO for road damage detection, they often rely on publicly available datasets, which may not represent local road conditions. Furthermore, the use of more advanced models like YOLOv9, which offer enhanced detection capabilities, has not yet been explored extensively in this context. This presents a significant research gap in terms of both model effectiveness and dataset relevance. To address this, the present study applies the YOLOv9 model to a custom-collected local dataset, aiming to provide more accurate and context-aware detection of various types of road damage.

II. MATERIALS AND METHOD

This research is designed to assess the performance of a road defect detection model using the YOLOv9 framework, a Convolutional Neural Network (CNN) architecture. After the problem is obtained and analyzed, a literature study is carried out to obtain references and relevant literature in solving the problem. This literature study is carried out by looking for references in the form of books or journals, either through internet media or other sources related to the use of Convolutional Neural Networks (CNN) and YOLOv9 in road damage detection. The literature found will provide a theoretical foundation and practical guidance in designing and implementing road defect detection models. This study adopts a systematic methodology consisting of five primary stages: data collection, data preprocessing, model training, model evaluation, and result analysis. Each stage plays a crucial role in building a robust road damage detection system using the YOLOv9 architecture. The overall research workflow is illustrated in Figure 1, which ensures the logical flow of the process from data acquisition to final evaluation.

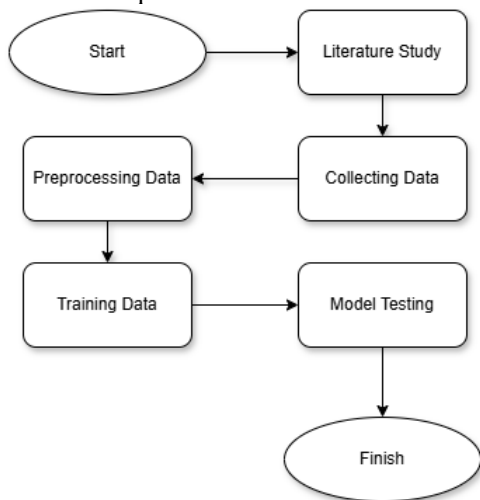


Fig. 1. Research Methods

Figure 1 presents the sequence of research stages: (1) Data Collection involves capturing high-resolution road images using a smartphone camera; (2) Data Preprocessing includes annotation, resizing, grayscale conversion, contrast adjustment, and augmentation; (3) Model Training utilizes YOLOv9 for learning features of road damage; (4) Model Evaluation measures performance metrics such as precision, recall, and mAP; and (5) Result Analysis interprets detection outcomes and training effectiveness.

A. Collecting Data

The dataset used in this research consists of road images that show various types of damage, such as cracks and potholes. The image capture process is done using a Techno brand smartphone camera device with a camera resolution of 108 Megapixels to ensure that the details of road damage can be seen clearly. The camera was used from a 45-degree viewing angle with a distance of one meter per object to optimally capture road conditions.

The research was conducted in Karawang Regency with data collection conducted directly in the field. Images were taken of paved and concrete roads in dry conditions to ensure the imaging results were not disturbed by environmental factors, such as water or reflection. The data collection locations included different types of roads, such as main roads, neighborhood roads, and industrial areas. As part of the dataset collection, several images were captured to represent different types of road damage. One common form is the occurrence of surface cracks, which are often caused by thermal expansion, excessive vehicle load, or degradation of road materials. These cracks vary in size and pattern, making them a challenge for consistent detection through conventional methods. An example of such damage is illustrated in Figure 2. This variation aims to cover the different types of road damage that may arise due to differences in environmental conditions and road usage levels.



Fig. 2. Cracked Road Image

Figure 2 shows a road segment with visible surface cracks, which appear as linear fractures along the pavement. The cracks typically emerge due to environmental stress, repeated traffic loads, or poor construction quality. Such defects are often subtle, and their detection requires high-resolution imagery

combined with advanced deep learning models for accurate classification. Another prevalent type of road damage captured in the dataset is potholes. These defects typically form due to the weakening of pavement structure from prolonged water infiltration, repeated traffic loads, or insufficient maintenance. Potholes pose a significant safety risk to road users, especially two-wheeled vehicles, due to the sudden depressions they create on the road surface. A visual example of this type of damage is presented in Figure 3.



Fig. 3. Pothole Road Image

This figure depicts a road segment containing one or more potholes—circular or irregular depressions on the pavement surface. These occur when water seeps into cracks, softens the underlying soil, and leads to collapse under the weight of vehicles. Potholes are generally easier to detect than surface cracks due to their distinct contours and depth variation, making them suitable for detection using object recognition models such as YOLOv9.

B. Preprocessing Data

This is an important step in data processing before further analysis or machine learning model training. In this research, the data preprocessing stage includes annotation, auto-orient, resize, grayscale, auto-adjust contrast and then perform augmentation to multiply the data. Annotation in this study using the Roboflow.com website is done manually on all images in the dataset using a bounding box to surround objects with low precision. Auto-orientation ensures consistent image orientation [17] to improve the accuracy of road damage detection.

Resize the process of adjusting the image size to fit the predefined dimensions [18], which aims to ensure size consistency and reduce computational complexity in image processing. Grayscale is a step that converts a color image into a gray-scale image. At this stage, all existing color information, such as differences in red, green, and blue hues, are completely removed. This removal of color information aims to make the model focus more on the important elements in the image, such as the visual structure, texture patterns, and shapes of objects, such as cracks and potholes on the road surface, without being affected by color variations that are irrelevant to the purpose of detection.

Next, the auto-adjust contrast process is performed using the adaptive histogram equalization technique. This method serves to improve image contrast by locally adjusting the pixel intensity distribution in the image over a small area. With sharper contrast, fine details such as crack edges, break lines, and hole contours become more visible and easily recognized by the detection model.

Augmentation is used to expand the amount and variety of training data in machine learning and computer vision. By modifying the original data through various techniques, augmentation aims to increase the diversity and quality of the training data, so that the model can learn more effectively and be able to deal with future data variations. In the data augmentation stage, we apply several techniques to expand the abundance of training data and improve model robustness.

The augmentation techniques applied include Rotating the image horizontally and vertically by flipping the image from left to right or from top to bottom, to simulate variations in the position of objects in the image. Rotation by 90 degrees or within a certain range: Rotate the image in a fixed angle (90°) or random angle so that the model can recognize objects even in different orientations. linear shift of the image.

Exposure level adjustment or changing the brightness of the image to create varying lighting conditions such as daylight, cloudy, or shadowy conditions. Noise augmentation by adding visual noise such as random spots or grain to train the model to remain robust despite image noise or imperfections. The application of these various augmentation techniques aims to enrich the training dataset and improve the model's ability to deal with the variety of situations that may occur in road damage images.

To build an effective YOLO configuration, a predetermined portion of test and trial data is necessary. In order to produce an accurate YOLO model, the initial data must be labeled and tested. The dataset consists of 1000 road images and is divided into three parts: 70% (700 images) for training, 10% (100 images) for validation, and 20% (200 images) for testing. This division of the dataset ensures the model can be properly trained, tested to measure its performance, and validated to avoid overfitting.

C. Yolo v9

YOLO (You Only Look Once) combines various object detection capabilities into a single neural network [19]. YOLO predicts each annotation box using the overall image features [20]. YOLO predicts that each bounding box for all object classes will open simultaneously in an image [21]. This shows that YOLO considers all aspects of the image, including all objects in it. In general, the YOLOv9 architecture consists of several main components:

1. **Backbone:** This part is responsible for extracting important features from the input image. YOLOv9 uses a more efficient and robust backbone than its predecessor, allowing the model to capture richer details.
2. **Neck:** This component serves to combine the features extracted by the backbone. By combining features from different levels, the model can obtain a more

comprehensive representation of the objects in the image.

3. Head: This part is responsible for performing bounding box and object class prediction. YOLOv9 uses a more sophisticated prediction mechanism, allowing the model to detect objects more accurately, especially small and adjacent objects.

The main innovation developed by YOLOv9 is GELAN (General Efficient Layer Aggregation Network) This new architecture is designed to maximize accuracy while minimizing the number of parameters and FLOPs (floating-point operations). GELAN allows the model to more efficiently process visual information and detect small objects.

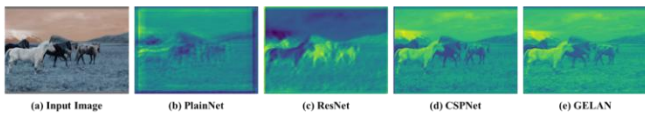


Fig. 4. How Yolov9 Works

Figure 4, sourced from [22], presents the visualization results of the output feature maps generated by random initial weights across various network architectures. The visualized data includes: (a) the input image, (b) PlainNet, (c) ResNet, (d) CSPNet, and (e) the proposed GELAN. Upon analyzing this figure, it becomes evident that, for each of the tested architectures, the information transmitted to the objective function for loss calculation is lost at varying rates. Specifically, the proposed architecture (GELAN) stands out by successfully retaining the most comprehensive information, while also providing the most reliable gradient information for the calculation of the objective function. This result highlights the superior capacity of GELAN to preserve and process relevant data through the network, offering a more robust approach to training deep learning models.

III. RESULT AND DISCUSSION

The results of the image training and validation process have been successfully formed into a comprehensive model. The primary theme of this paper centers around the findings and advancements in the field of road damage detection. In order to thoroughly analyze and evaluate real-world road damage, we opted to utilize manual data collection as a part of the research scenario, as opposed to relying on publicly available data. This decision was made deliberately, as the database used in the manual data collection process is more representative of the specific context and scope of the research addressed in this paper. By using this tailored data collection approach, we ensure that the model is better aligned with the practical aspects of road damage detection, offering insights that are more relevant and applicable to real-life scenarios.

The YOLO modeling data was tested to detect objects in images. To assess the accuracy of the object data classifier, calculations are performed using packages such as matplotlib, numpy, sklearn, and torch. The training results from these packages will be displayed in the form of diagrams and images, which will facilitate data analysis. In the test, the epoch used was 100; with a training batch size of 16, which means 16 images were processed at once in one iteration. In addition, the

image size used was 640 x 640 pixels, and all training was conducted using Google Collaboratory. Table 1 shows the detection results of cracked and potholed roads.

TABLE I. DETECTION RESULTS

Image	Damage type
	Pothole
	Crack

After detection testing, road defects with a confidence of 0.52 were detected in the pothole type, while details of small cracks at the end of the road could be detected with a confidence of 0.54. Although the confidence level obtained is quite low, the faint objects can still be detected well. On the other hand, objects that are more clearly visible to the naked eye show a higher confidence level, especially for potholes with a confidence level of 0.95. For cracked road damage, the confidence level recorded was 0.91. The resulting F1-Score graph shows that potholes are easier to detect than cracks. This can be seen in Figure 5, which displays the F1-Score curve.

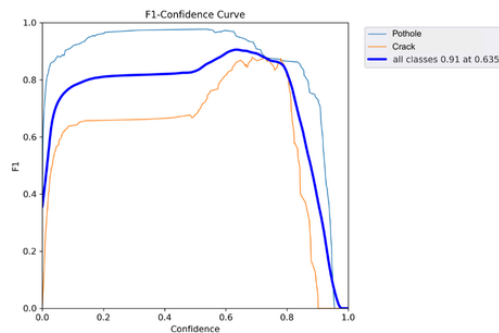


Fig. 5. F1-Score Curve

Meanwhile, the results of the Precision Confidence measurement showed that all types of road defects, without

exception, achieved a score of 0.85. This reflects a fairly high level of accuracy in road defect detection. This result can be clearly seen in the curve depicted in Figure 6, which presents a visualization of the comparison of the detection confidence level with the achieved precision. The curve provides a more complete picture of the model's accuracy in detecting road defects, which further strengthens the validity of the findings in this study.

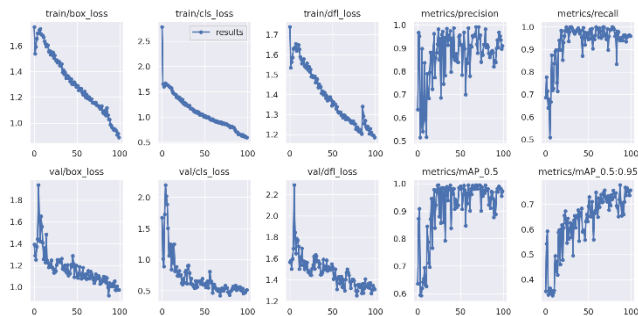


Fig. 6. Loss Chart

The Precision Recall result for pothole road damage is 0.992, and for cracked roads it reaches 0.94, which can be seen in Figure 7.

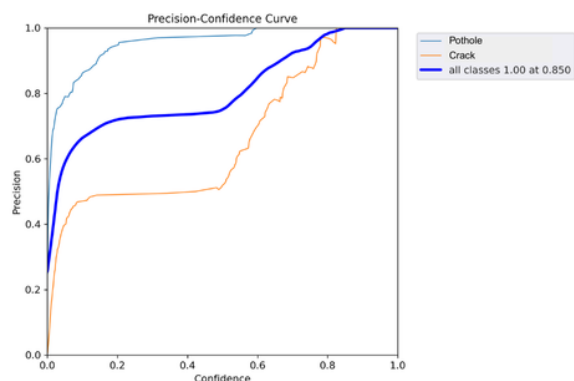


Fig. 7. Precision Confidence Curve

In addition to displaying the Precision results, in this study we also present graphs depicting the loss values during the training and testing process. These graphs include various metrics, ranging from train loss to validation loss, which provide a detailed picture of the model's performance throughout the training process. This process is important to evaluate how well the model can generalize data that was not seen before, as well as to identify potential overfitting or underfitting. These results can be clearly seen in Figure 8, which shows a comparison between train loss and validation loss at various epochs, providing further insight into the effectiveness of our model training and validation process.

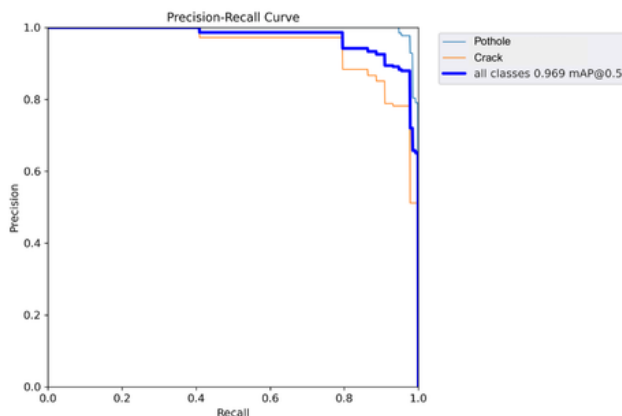


Fig. 8. Precision Recall Curve

In Figure 8 the training result graph displays the box_loss, cls_loss, and dfl_loss metrics for the train and val data. These three losses show a decrease at each epoch, indicating that the model is able to learn the task well. Box_loss measures the fit of the bounding box, cls_loss is related to object detection, and dfl_loss indicates classification accuracy. The val graph shows larger fluctuations than the train due to the difference in the amount of data. In addition, the precision, recall, mAP50, and mAP50-95 metrics graphs show an increasing trend every

epoch, reflecting the improvement in classification accuracy and performance. The mAP50 value reaches 0.96, indicating excellent detection accuracy, while mAP50-95 reaches 0.77, indicating a low error rate and good classification performance.

IV. RESULT AND DISCUSSION

The YOLOv9 model demonstrated strong performance in detecting road defects, specifically cracks and potholes. The model achieved a precision of 0.85 and a recall of 0.992 for pothole detection, and a precision of 0.94 for crack detection, indicating its ability to accurately identify and classify these defects. Furthermore, the mAP50 of 0.96 and mAP50-95 of 0.77 further support the model's high detection and classification accuracy. These results demonstrate the model's effectiveness in handling variations in detection difficulty across different defect types.

Furthermore, the model exhibited a consistent decrease in the loss function value throughout the training process. This decline indicates effective learning and optimization of the model's performance during training. The successful implementation of YOLOv9 for road defect detection demonstrates its potential as a foundation for developing an automated road monitoring system. Such a system could significantly aid relevant authorities in expediting the identification and remediation of road defects, thereby enhancing road user safety and improving the efficiency of road infrastructure maintenance.

REFERENCES

- [1] T. Ngabalin, A. F. Habibie, and E. Darmawan, "Strategi Pengembangan Ekosistem Ekonomi Kreatif Dalam Mendukung Kebijakan Pariwisata Di Kota Tanjungpinang," *J. Ilmu Adm. Negara*, vol. 11, no. 01, pp. 13–21, 2023, doi: 10.31629/juan.v11i01.5839.
- [2] R. Der Sarkissian, C. Abdallah, J. M. Zaninetti, and S. Najem, "Modelling intra-dependencies to assess road network resilience to natural hazards," *Nat. Hazards*, vol. 103, no. 1, 2020, doi: 10.1007/s11069-020-03962-5.
- [3] M. T. A. Siregar, "Upaya Yang Dapat Dilakukan Oleh Korban/ Pengguna Jalan Meminta Pertanggungjawaban Pidana Penyelenggara Jalan Atas Terjadinya Kecelakaan Akibat Jalan Rusak," *EduTech J. Ilmu Pendidik. dan Ilmu Sos.*, vol. 6, no. 1, pp. 37–45, 2020, doi: 10.30596/edutech.v6i1.4393.
- [4] F. B. Santosa and A. S. Amal, "Kajian Studi Kelayakan Rencana Pembangunan Jalan Srdondol-Sekaran Kota Semarang," *Semin. Keinsinyuran Progr. Stud. Progr. Profesi Ins.*, vol. 3, no. 1, 2023, doi: 10.22219/skpsppi.v3i1.7544.
- [5] R. Rivaldo and F. R. Yamali, "Perencanaan Perkerasan Kaku (Rigid Pavement) Ruas Jalan Hitam Ulu-Mentawak Di Kabupaten Merangin (Menggunakan Metode AASHTO 1993)," *J. Talent. Sipil*, vol. 5, no. 1, 2022, doi: 10.33087/talentsipil.v5i1.95.
- [6] P. Toscani, W. Sekot, and F. Holzleitner, "Forest roads from the perspective of managerial accounting-empirical evidence from Austria," *Forests*, vol. 11, no. 4, 2020, doi: 10.3390/F11040378.
- [7] N. I. Sembiring, R. Siahaan, and P. Naibaho, "Analysis of Damage Conditions on the Berastagi-Simpang Empat Road, Karo Regency, Using PCI and Sdi Methods," *Tjybjb.Ac.Cn*, vol. 3, no. 2, pp. 58–66, 2021, [Online]. Available: <http://www.tjybjb.ac.cn/CN/article/downloadArticleFile.do?attachType=PDF&id=9987>
- [8] S. Suryadi and F. A. P. Nasution, "Revolusi Industri, Tren Pekerjaan Masa Depan, dan Posisi Indonesia," *J. Ketenagakerjaan*, vol. 18, no. 2, pp. 124–141, 2023, doi: 10.47198/jnaker.v18i2.237.
- [9] S. Kiranyaz, O. Avci, O. Abdeljaber, T. Ince, M. Gabbouj, and D. J. Inman, "1D convolutional neural networks and applications: A survey," *Mech. Syst. Signal Process.*, vol. 151, 2021, doi: 10.1016/j.ymssp.2020.107398.
- [10] Q. Chen, X. Gan, W. Huang, J. Feng, and H. Shim, "Road damage detection and classification using mask R-CNN with DenseNet backbone," *Comput. Mater. Contin.*, vol. 65, no. 3, 2020, doi: 10.32604/cmc.2020.011191.
- [11] Y. Pan, X. Chen, Q. Sun, and X. Zhang, "Monitoring Asphalt Pavement Aging and Damage Conditions from Low-Altitude UAV Imagery Based on a CNN Approach," *Can. J. Remote Sens.*, vol. 47, no. 3, 2021, doi: 10.1080/07038992.2020.1870217.
- [12] A. Antoni, T. Rohana, and A. R. Pratama, "Implementasi Algoritma Convolutional Neural Network Untuk Klasifikasi Citra Kemasan Kardus Defect dan No Defect," *Build. Informatics, Technol. Sci.*, vol. 4, no. 4, pp. 1941–1950, 2023, doi: 10.47065/bits.v4i4.3270.
- [13] R. Yati, T. Rohana, and A. R. Pratama, "Klasifikasi Jenis Mangga Menggunakan Algoritma Convolutional Neural Network," *J. Media Inform. Budidarma*, vol. 7, no. 3, p. 1265, 2023, doi: 10.30865/mib.v7i3.6445.
- [14] A. Kirana, H. Hikmayanti, and J. Indra, "Pengenalan Pola Aksara Sunda dengan Metode Convolutional Neural Network," *Sci. Student J. Information, Technol. Sci.*, vol. 1, no. 2, pp. 95–100, 2020, [Online]. Available: <http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/download/19/15>
- [15] K. Ahmad Baihaqi and C. Zonyfar, "Deteksi Lahan Pertanian Yang Terdampak Hama Tikus Menggunakan Yolo v5," *Syntax J. Inform.*, vol. 11, no. 02, pp. 01–11, 2022, doi: 10.35706/syji.v11i02.7226.
- [16] S. Ma'arif, T. Rohana, and K. Ahmad Baihaqi, "Deteksi Jenis Beras Menggunakan Algoritma YOLOv3," *Sci. Student J. Information, Technol. Sci.*, vol. III, pp. 219–226, 2022, [Online]. Available: <http://journal.ubpkarawang.ac.id/mahasiswa/index.php/ssj/article/view/443/357>
- [17] S. Informasi, F. N. Fajri, and W. G. Priambodo, "Jurnal Advance Research Informatika Fire and Smoke Object Detection Using Mask R-CNN," vol. 2, no. 2, pp. 1–7, 2024.
- [18] V. R. Atvira, "Deteksi Kualitas Kemurnian Susu Sapi Melalui Pengolahan Citra Digital Menggunakan Metode Active Contour Dengan Klasifikasi K-Nearest Neighbor," *Univ. Telkom Bandung*, vol. 6, no. 2, pp. 3830–3837, 2019, [Online]. Available: <https://openlibrary.telkomuniversity.ac.id/pustaka/152822/>
- [19] U. Rahardja, M. A. Ngad, S. Millah, E. P. Harahap, and Q. Aini, "Blockchain Application in Educational Certificates and Verification Compliant with General Data Protection Regulations," in *2022 10th International Conference on Cyber and IT Service Management, CITSM 2022*, 2022, doi: 10.1109/CITSM56380.2022.9935909.
- [20] N. Lutfiani, S. Wijono, U. Rahardja, A. Iriani, and E. A. Nabila, "Artificial Intelligence Based on Recommendation System for Startup Matchmaking Platform," in *2022 IEEE Creative Communication and Innovative Technology, ICCIT 2022*, 2022, doi: 10.1109/ICCIT55355.2022.10118708.
- [21] T. Hongsuchon *et al.*, "Brand Experience on Brand Attachment: The Role of Interpersonal Interaction, Feedback, and Advocacy," *Emerg. Sci. J.*, vol. 7, no. 4, 2023, doi: 10.28991/ESJ-2023-07-04-014.
- [22] C. Wang, I. Yeh, and H. M. Liao, "Supplementary of YOLOv9," pp. 1–7.