Telematika: Jurnal Informatika dan Teknologi Informasi *ISSN: 1829-667X / E-ISSN: 2460-9021*
Vol. 22, No. 1, Februari 2025, pp.65-82 *DOI:10.31515/telematika.v22i1.14802*

Improving the Efficiency of Water Meter Reading at Perumdam Tirta Kerta Raharja Using Microcontroller-Based Implementation of the YOLOv9 Method

Peningkatan Efisiensi Pembacaan Angka Meter Air Perumdam Tirta Kerta Raharja Berbasis Mikrokontroler dengan Penerapan Metode Yolov9

Septian Ade Putra[1], Nita Merlina[2]

[1,2] Ilmu Komputer, Universitas Nusa Mandiri Margonda, Indonesia

[1*]14220026@nusamandiri.ac.id, [2]nita@nusamandiri.ac.id
**: Penulis korenspondensi* (*corresponding author*)

Informasi Artikel
Received: December 2024
Revised: January 2025
Accepted: January 2025
Published: February 2025

Abstract

Manual water meter reading remains a challenge for Perumdam Tirta Kerta Raharja due to its labor-intensive process, susceptibility to human errors, and inefficiency. This study aims to develop an automated water meter reading system using YOLOv9 and a microcontroller to improve efficiency and data accuracy. The model was trained using a dataset of water meter images under various lighting conditions and viewing angles. Evaluation results indicate that the 20-epoch configuration is the best model, achieving 99,91% accuracy, 91,16% average precision, and 91,04% average recall. The developed system successfully detects digits in real-time with high accuracy when deployed on a Raspberry Pi-based platform. However, the model still faces challenges in detecting the Background class. With further optimization, this system can be widely implemented to enhance operational efficiency in Perumdam and related industries.

Keywords: Water meter reading detection; YOLOv9; Epoch configuration; Microcontroller
Kata kunci: Deteksi angka meter air; Yolov9; Konfigurasi epoch; Mikrokontroler

Abstrak

Pembacaan angka meter air secara manual masih menjadi tantangan bagi Perumdam Tirta Kerta Raharja karena prosesnya yang membutuhkan banyak tenaga kerja, rentan terhadap kesalahan manusia, dan kurang efisien. Penelitian ini bertujuan mengembangkan sistem otomatisasi pembacaan angka meter air berbasis YOLOv9 dengan mikrokontroler untuk meningkatkan efisiensi dan akurasi pencatatan data pelanggan. Model dilatih menggunakan dataset gambar meter air dalam berbagai kondisi pencahayaan dan sudut pengambilan gambar. Hasil evaluasi menunjukkan bahwa konfigurasi 20 epoch merupakan model terbaik, dengan akurasi 99,91%, presisi rata-rata

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

91,16%, dan recall rata-rata 91,04%. Sistem yang dikembangkan berhasil mendeteksi angka secara real-time dengan tingkat keberhasilan tinggi pada deployment berbasis Raspberry Pi. Meskipun demikian, model masih mengalami kesulitan dalam mendeteksi kelas Background. Dengan optimasi lebih lanjut, sistem ini dapat diterapkan secara luas untuk meningkatkan efisiensi operasional Perumdam dan industri terkait.

1. Introduction

Efficient and accurate water meter recording is a major challenge for Perumdam Tirta Kerta Raharja, especially with the increasing number of customers relying on manual methods that are slow, error-prone, and inefficient. Advances in technology, such as computer vision and object detection algorithms like YOLOv9, offer solutions for automating water meter readings with high accuracy [1]. This study aims to develop a YOLOv9 model capable of converting numeric images on water meters into real-time numerical data, thereby improving operational efficiency, reading accuracy, and supporting better water resource management.

The research adopts a transfer learning approach to train the YOLOv9 model using an annotated dataset of water meter numeric images. This technology is integrated with a camera-enabled microcontroller connected via Wi-Fi, enabling automatic readings under various lighting conditions and device configurations. Expected outcomes include reduced human error, cost efficiency, and enhanced customer service through a reliable automated system. Additionally, this study involves the development of a customized dataset and data augmentation techniques to address operational challenges, offering practical and economical solutions that can be broadly implemented by other water utilities.

1.1. Image Processing Classification

Classification is one of the most commonly used techniques in data mining, consisting of two main stages: learning and classification. In the learning stage, training data is analyzed using a classification algorithm to recognize patterns or groups within the data. The primary goal of classification is to group data based on specific patterns found, so the entire dataset can be categorized into certain classes based on target attributes or outputs. This approach is highly useful for identifying behavioral patterns in data and serves as a data discrimination mechanism [2].

In the field of image processing and computer vision, classification aims to group images or parts of images based on specific features extracted from the image. The image processing classification process involves five main stages. First, the preprocessing stage is performed to improve image quality, such as noise removal or contrast adjustment, making it easier to extract the desired features. Second, the feature extraction stage aims to identify key features from the image, such as edges, textures, or shapes, which form the basis for recognizing patterns. Third, dimensionality reduction reduces data complexity while retaining critical features, enabling more efficient processing without losing crucial information. Fourth, the learning stage involves training the classification model using training data to learn patterns from the extracted features, where learning algorithms are applied to build the model. Finally, in the classification stage,

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

new images are tested using the trained model to determine the appropriate class or category based on the recognized features [3] [4].

1.2. You Only Look Once (YOLO)

YOLO (You Only Look Once) is a deep learning algorithm based on a convolutional neural network (CNN) used for object detection, target identification, and localization. The YOLO process consists of three main stages: preprocessing, processing, and classification [5]. This technology continues to evolve to meet real-time performance demands with high accuracy, particularly in industrial sectors like surface detection. In its latest version, YOLOv9 introduces innovations such as Programmable Gradient Information (PGI) to address information loss during training and the Generalized Efficient Layer Aggregation Network (GELAN) to enhance computational efficiency. These advancements make YOLOv9 superior for real-time object detection with high accuracy and speed, making it relevant for various applications in dynamic environments [6] [7].

The Information Bottleneck principle aims to extract relevant information from input data required for output prediction while minimizing irrelevant information. This principle focuses on the trade-off between data compression and information relevance, using mutual information to measure retained information. This concept is applied in machine learning to understand neural network internal representations and improve data compression efficiency. Technologies like YOLO leverage this principle to operate in real-time with high efficiency, enabling accurate object detection and classification in images [8].

YOLO can be applied to automate water meter readings in urban environments. This technology allows systems to detect and interpret numbers on meters accurately and in real-time, replacing manual methods prone to errors. This implementation improves the efficiency and speed of readings, making it a superior solution for industrial applications [9].

This study employs the YOLOv9 method as a CNN architecture to detect and read water meter numbers in real-time. The process involves collecting and labeling image datasets, training the YOLOv9 model, and evaluating its performance using accuracy and precision metrics. Image preprocessing techniques are applied to prepare images for training, while data augmentation is used to enhance dataset size and diversity.

The YOLOv9 model is tailored for water meter reading tasks using the YOLOv9 architecture, as shown in Figure 1. This process involves modifying the model architecture to improve accuracy under diverse conditions and training the model using the collected dataset, with data augmentation techniques applied to enrich the training data's variation.

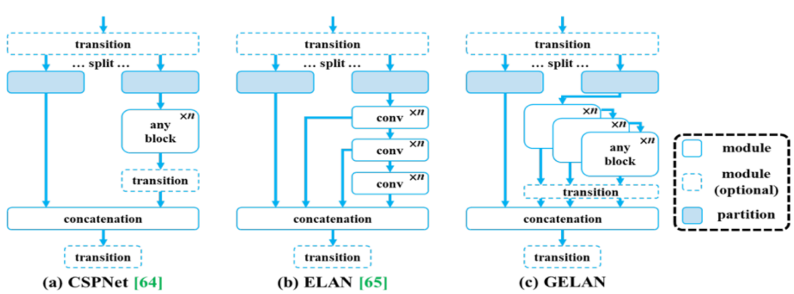


Figure 1. GELAN Architecture in YOLOv9 [10]

YOLOv9 introduces several innovations in network architecture, such as enhancements in spatial pyramid pooling (SPP), efficient layer aggregation network (ELAN), and optimized Cross Stage Partial Network (CSPNet). This combination enables YOLOv9 to detect objects more quickly and accurately in high-resolution images, which are commonly encountered in the process of capturing water meter images [10].

1.3. Model Evaluation

Model evaluation aims to measure the performance of a machine learning or deep learning model in predicting and completing specific tasks. In YOLO (You Only Look Once), evaluation utilizes a confusion matrix to compare the model's prediction results with the ground truth data [11]. The confusion matrix consists of four main components: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN). TP indicates correctly detected objects, FP represents detection errors, and FN refers to undetected objects [12].

Evaluation metrics such as Precision, Recall, F1-Score, and Accuracy are used to analyze the model's performance, as shown in equations (1) through (4) [13].

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1 - Score = 2\ x\ \frac{Recall\ x\ Precission}{Recall+Precission} \quad (4)$$

The confusion matrix is also used to calculate the mean Average Precision (mAP), which serves as the standard for evaluating YOLO models. mAP calculates the average precision for each detected class, as represented in Equation (5) [14].

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \quad (5)$$

Where n s the number of classes, and AP_k represents the Average Precision for the k-th class. This evaluation balances accuracy and speed, two critical aspects of YOLO's performance.

1.4. Cross-Industry Standard Process for Data Mining (CRISP-DM)

CRISP-DM (Cross-Industry Standard Process for Data Mining) is a widely used framework in data mining projects across various industries. Developed in the late 1990s by companies such as IBM and NCR, CRISP-DM provides an iterative, flexible, and process-based approach to uncovering patterns and insights from data to support business decision-making. This framework consists of six main stages. The first stage is Business Understanding, which focuses on understanding business objectives, planning the analysis, and determining the metrics to be measured. Next, Data Understanding involves collecting and exploring data to understand its characteristics and identify data quality issues. The third stage is Data Preparation, which involves cleaning, transforming, and organizing the data to make it ready for use in analytical models. Then, Modeling is conducted by building predictive models using techniques such as decision trees or neural networks. The fifth stage is Evaluation, where the model results are assessed to ensure alignment with business objectives. Finally, Deployment involves implementing the model into operational systems and business processes [15]. Although CRISP-DM's flexibility, iteration, and focus on business objectives are key advantages, the

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

framework also has limitations, such as the time-consuming data preparation process and challenges in handling projects with very large data volumes. The stages of the CRISP-DM method are shown in Figure 2.

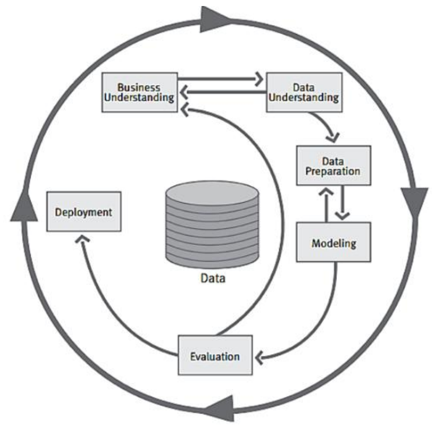


Figure 2. CRISP-DM Stages [16]

1.5. Previous Research

Previous studies have discussed the Automatic Meter Reading (AMR) approach using a Convolutional Neural Networks (CNN) architecture in two stages: Fast-YOLO for meter reading location detection and the CNN model for meter number recognition. The main contribution was the introduction of the UFPR-AMR dataset, which consists of 2,000 fully annotated images and is the largest public dataset in AMR literature. Data augmentation techniques were applied to improve model performance, and the results showed high performance despite using a small dataset, with a trade-off between speed and accuracy [17]. Another study focused on automatic dial meter reading using the UFPR-ADMR dataset, a real-world dataset for dial meters. This study achieved a 100% F1-score in dial detection using Faster R-CNN and YOLO, and recognition rates of 93.6% for the dial and 75.25% for the meter with Faster R-CNN (ResNext-101) [18].

Research on automatic water meter reading using deep learning also divided the task into several subtasks, such as component localization and orientation alignment, to address environmental challenges. The results demonstrated the effectiveness of this method, even under varying lighting conditions and obstructions [19]. Additionally, research on construction workers' helmet detection using YOLOv2 showed success in detecting workers wearing or not wearing helmets, with an F1-score of 0.79 using the Non-Maximum Suppression algorithm to improve precision [20].

The final study discussed helmet usage detection for motorcyclists, where helmets were found to reduce the risk of death by 40% and severe injuries by over 70%. Using CNN-based computer vision technology, the system was able to detect motorcyclists without helmets with 90% accuracy, which could help reduce traffic accidents and facilitate police surveillance tasks [21].

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

2. Methode

This study combines statistical analysis techniques to measure model performance and qualitative methods to assess technology acceptance by Perumdam staff. The objective is to provide a comprehensive overview of the benefits and challenges of implementing the YOLOv9 model. An experimental research design with a quantitative approach is used to test the effectiveness of the YOLOv9 algorithm in the water meter reading system, which is conducted in two stages: model development and model validation in an operational environment.

2.1. Data Collection

The research data consists of water meter digit images obtained from Perumdam Tirta Kerta Raharja. Data collection involves capturing images of water meters under various lighting conditions and angles to train the model with diverse environmental variations.

2.2. Data Pre-Processing

The pre-processing stage includes image annotation, data augmentation, and the use of FiftyOne. Image annotation is performed manually using the Roboflow tool [22], where each image is labeled with bounding boxes around essential elements. Data augmentation is applied to training images to increase variation, using techniques such as rotation, flipping, lighting adjustments, and zooming without modifying annotations [23]. The FiftyOne process is utilized for dataset management, visualization, and analysis to ensure a balanced class distribution and consistent annotations, facilitating dataset quality validation and YOLO model evaluation.

2.3. Modeling

The modeling process involves developing and adapting the YOLOv9 object detection model to recognize digits on water meters at Perumdam Tirta Kerta Raharja. The process begins with optimizing the YOLOv9 architecture to accommodate operational conditions such as lighting and meter positioning. The model is trained using the collected dataset, with data augmentation techniques applied to enhance robustness. Validation is conducted in an operational environment to assess the accuracy and speed of the reading process. The dataset is split into 70% for training, 15% for validation, and 15% for testing. Performance evaluation includes metrics such as mAP (Mean Average Precision), precision, recall, F1-score, accuracy, and inference speed to ensure the model's reliability for real-world applications.

2.4. Analysis and Evaluation

A comprehensive approach to water meter reading evaluation includes a comparative assessment of object detection algorithms, such as YOLOv9, to determine the most effective method for Perumdam Tirta Kerta Raharja. The primary focus of the YOLOv9 evaluation is on reading accuracy, processing speed, and error rates under varying operational conditions. The analysis employs performance metrics such as mAP, precision, recall, and F1-score to measure the model's ability to detect and classify objects correctly. Additionally, the evaluation examines prediction errors, such as false positives and false negatives, and assesses the model's robustness against variations in image conditions, including lighting and angle differences.

3. Results And Discussion

This study applies the YOLOv9 model to detect numbers on water meters with a configuration adjusted through several iterations to achieve optimal performance. Experiments were conducted with three configurations for the number of epochs (10, 20, and 50) to evaluate the

 ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

model's performance at different training levels. The hyperparameters used include a batch size of 8, an initial learning rate of 0.001, and an image size of 640. Training was performed using CUDA-enabled devices with 8 workers for efficient parallel processing. Data augmentation was enabled to increase data diversity, while the learning rate was adjusted using the Cosine Annealing Scheduler to maximize model convergence. This configuration aimed to achieve high accuracy in quickly detecting numbers. The YOLOv9e model used has 687 layers, 57.4 million parameters, and requires 189.2 GFLOPs for computation.

The dataset was split into three subsets for training, validation, and testing with a ratio of 70%:15%:15%. The data division was carried out incrementally, resulting in 761 training samples, 163 validation samples, and 164 test samples. The data distribution for labels 0 through 9 helps the YOLOv9 model learn detection patterns more effectively. The data distribution in FiftyOne is shown in Figure 3.

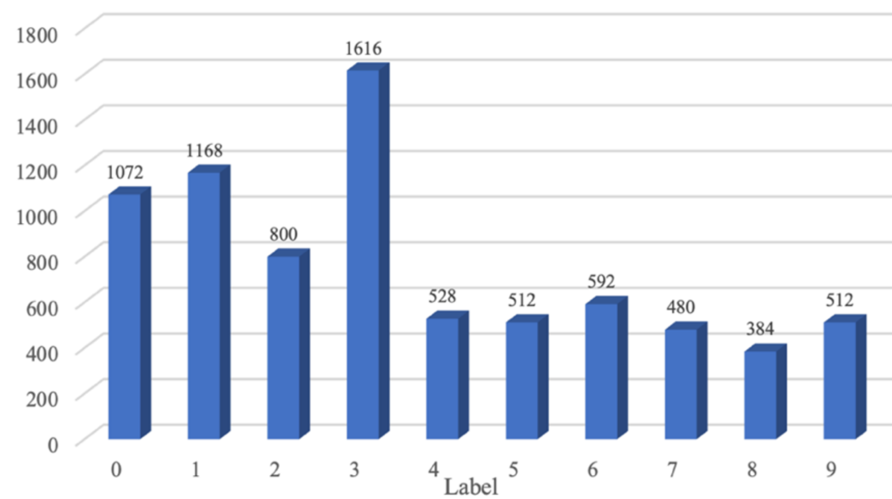


Figure 3. Distribution of Water Meter Number Labels

The model evaluation results after 10 epochs are presented in Table 1, with the overall model performance showing a precision of 99,90% and recall of 0.868.

The first configuration of the YOLOv9 model for training over 10 epochs resulted in performance evaluation displayed in a graph illustrating the relationship between object detection accuracy (metrics/mAP50 and metrics/mAP50-95) and the decrease in bounding box prediction errors (train/box_loss), as shown in Figure 4. The graph indicates a positive trend, where the loss value decreases as the number of epochs increases, while metrics/mAP50 and metrics/mAP50-95 improve, signifying an enhancement in the model's object detection capability.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

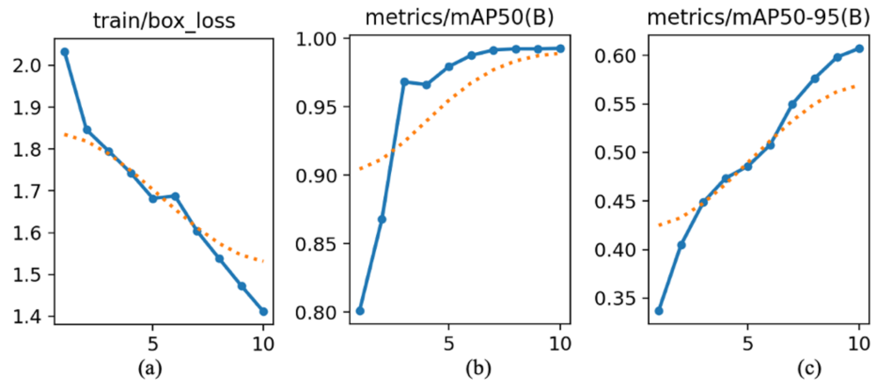


Figure 4. Model Performance During 10 Epoch Training Process
(a) Decrease in Bounding Box Prediction Error; (b-c) Object Detection Accuracy

Model evaluation after 10 epochs demonstrated excellent performance, as shown in Table 1.

Table 1. YOLOv9 Model Evaluation Results for 10 Epochs

Class	Image	Instances	Precision	Recall
All	163	1.151	0,990	0,994
0	96	142	0,997	0,993
1	126	180	0,965	0,994
2	87	127	1,000	1,000
3	161	238	0,988	1,000
4	64	81	0,995	1,000
5	67	84	1,000	0,987
6	67	87	0,974	0,989
7	62	73	0,999	0,973
8	50	57	0,995	1,000
9	68	82	0,982	1,000

The model was tested on 163 images with a total of 1,151 digit instances, achieving 99% precision and 99.4% recall. Some classes, such as class 2 and 4, exhibited perfect performance with 100% precision and recall. While most classes demonstrated high performance, some, like class 1 (precision 96.5%) and class 7 (recall 97.3%), had minor deficiencies in detection accuracy. These results indicate that the model has high accuracy in detecting digits on water meters, though there is still room for further optimization.

Confidence curve analysis shows a balance between precision and recall at various confidence thresholds, as depicted in Figure 5. The model maintains high F1-score, precision, and recall values up to a confidence threshold of approximately 0.8 before experiencing a decline. Precision remains stable near 1.0, while recall begins to drop as the confidence threshold increases. These results confirm the model's reliability in reading digits with high accuracy, though improvements can be made to maintain a balance between precision and recall under different conditions.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

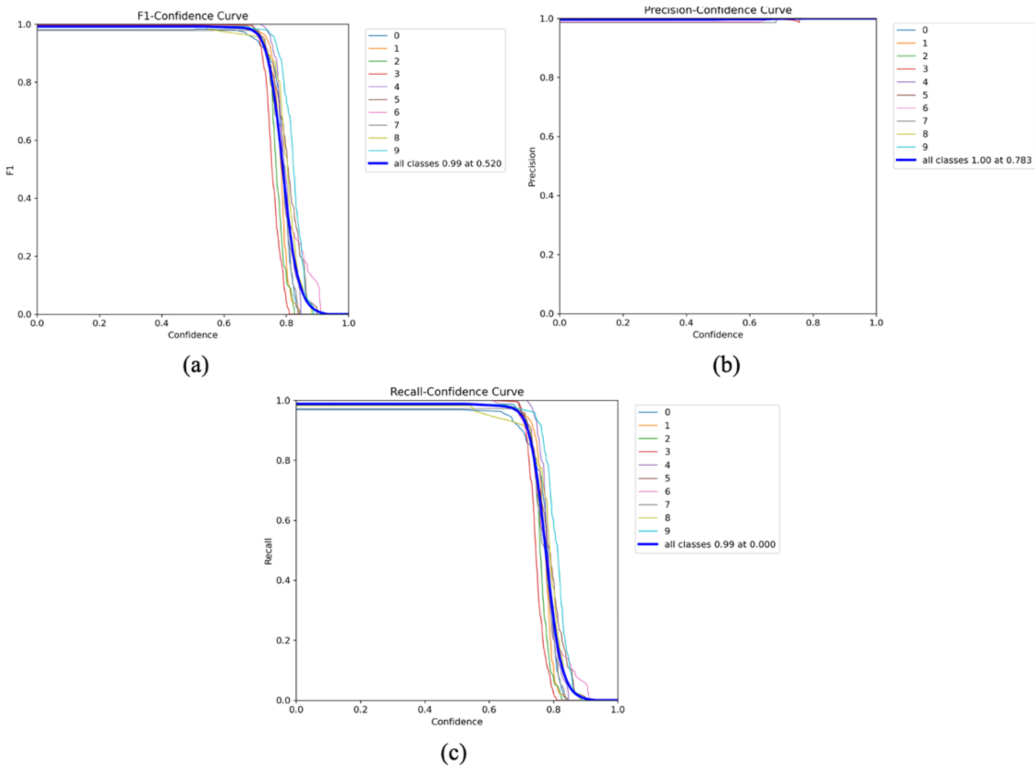
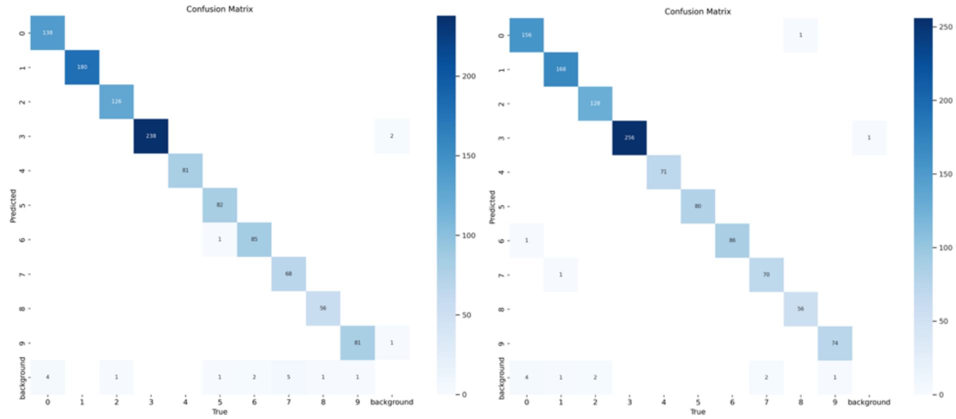


Figure 5. Confidence Curve During 10 Epoch Testing Process
(a) F1-Confidence Curve, (b) Precision-Confidence Curve, and (c) Recall-Confidence Curve

Evaluation using a confusion matrix reveals a very high accuracy level, as shown in Figure 6, with most classes having a high number of correct predictions. During validation, the confusion matrix shows that the model has a very high accuracy level, with most classes achieving precise predictions. Precision and recall for most classes approach 1.0, indicating the model's capability to accurately identify digits in validation data. However, minor misclassifications occur, particularly in classes with fewer samples. In testing, the confusion matrix highlights greater challenges in detecting digits in new images. Although overall accuracy remains high (99.78%), some classes experience a performance drop, particularly in distinguishing objects from the background. Misclassifications are more prominent in class 9 and the "Background" class, which have precision, recall, and F1-score of 0%, indicating that the model still struggles to identify certain elements outside the training data. This suggests that the model is better at recognizing patterns in validation data than in test data, necessitating further optimization, such as increasing training data or refining hyperparameters, to enhance the model's generalization to new data.

A comparison of accuracy, precision, recall, and F1-score for the validation and testing processes over 10 epochs is presented in Table 2.

Telematika: Jurnal Informatika dan Teknologi Informasi
Vol. 22, No. 1, Februari 2025, pp.65-82
ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802



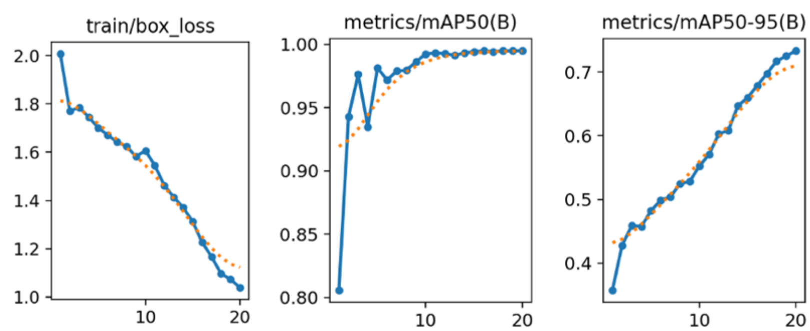
(a) (b)

Figure 6. Confusion Matrix for 10 Epochs; (a) Validation Process (b) Testing Process

Table 2. Comparison of Accuracy, Precision, Recall, and F1-score for 10 Epochs

Metric	Validation Process	Testing Process
Accuracy	99,78%,	99,63%
Precision	80,66%	89,70%
Recall	89,73%	89,60%
F1-Score	89,73%	89,70%

The YOLOv9 model trained for 20 epochs showed a significant performance improvement compared to the previous 10 epochs. A decrease in train/box_loss from approximately 2.0 to 1.0 indicates improved accuracy in detecting bounding boxes, as shown in Figure 7.



(a) (b) (c)

Figure 7. Model Performance During 20 Epoch Training Process
(a) Decrease in Bounding Box Prediction Error; (b-c) Object Detection Accuracy

Furthermore, the mAP@50 metric approached 1.0, and mAP@50-95 increased from 0.35 to over 0.75, reinforcing the model's superiority. Evaluation on 163 images with 1,151 instances showed that the model achieved a precision of 0.993 and a recall of 0.997, reflecting an excellent detection level. Most classes exhibited high precision and recall, with classes 5 and 7 reaching 1.000, while the lowest precision was found in class 8 at 0.983.

Confidence curve analysis revealed that the F1-Confidence Curve remained high across most confidence thresholds, though experiencing a decline near 1.0, as shown in Figure 8.

Telematika: Jurnal Informatika dan Teknologi Informasi
Vol. 22, No. 1, Februari 2025, pp.65-82

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

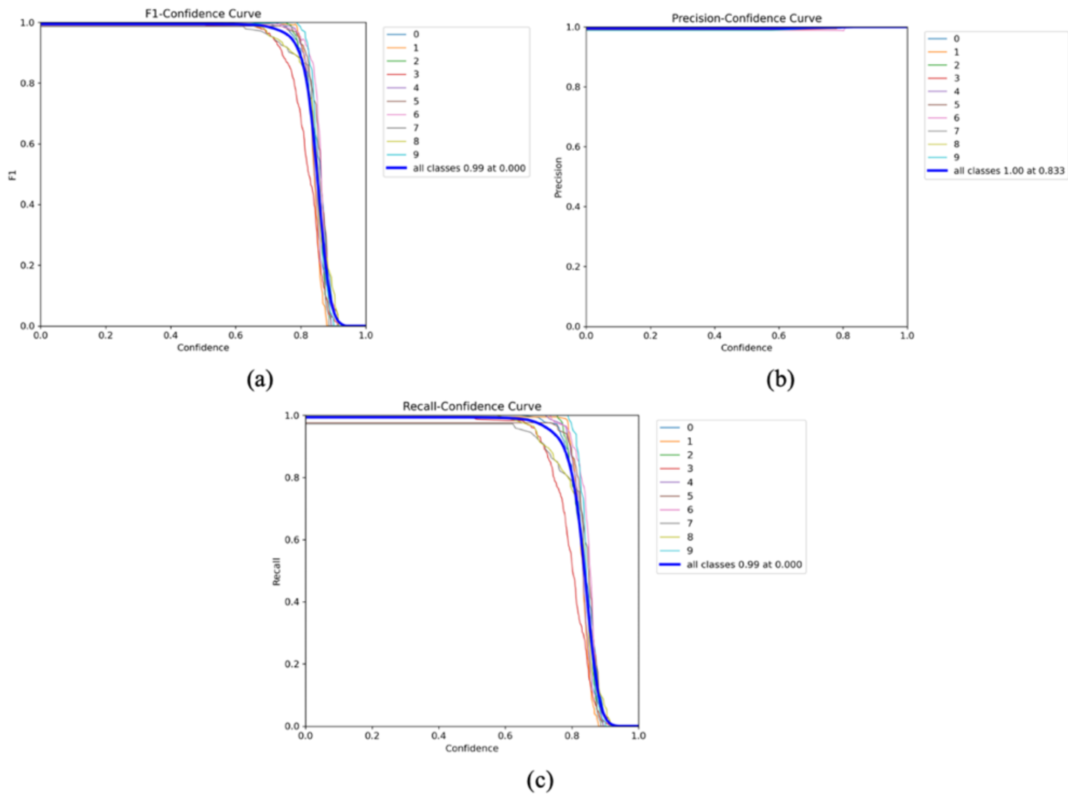


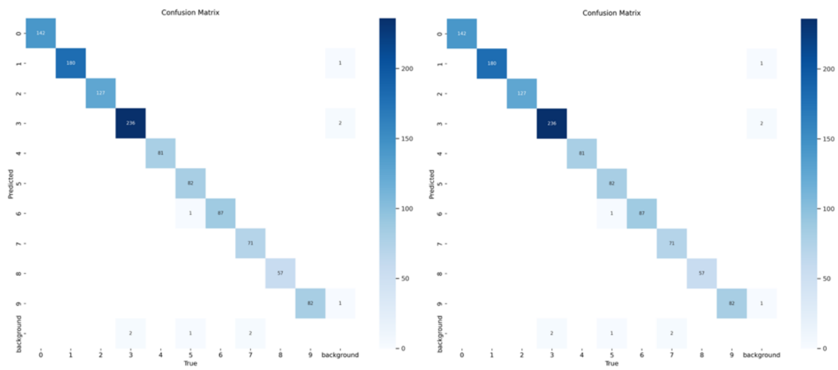
Figure 8. Confidence Curve During 20 Epoch Validation Process
(a) F1-Confidence Curve, (b) Precision-Confidence Curve, and (c) Recall-Confidence Curve

The Precision-Confidence Curve demonstrates high stability, with a slight decline after the confidence threshold reaches 0.78. Meanwhile, the Recall-Confidence Curve experiences a drop at very high confidence thresholds, indicating an increase in false negatives. During the validation process, the model exhibits an accuracy of 99.84%, with an average Precision of 90.7% and Recall of 90.6%. However, some classification errors occur, particularly in class 3 with two errors and in classes 6 and 9 with one error each. The model also struggles to detect the background class, with Precision and Recall values of 0.000.

In the testing process, a similar pattern is observed in the Confidence Curve, where Precision remains stable while Recall declines as the confidence threshold increases. The classification error rate is higher compared to validation, especially for certain classes, and difficulties in recognizing the background remain a challenge. Overall, the model demonstrates improved detection compared to the previous 10 epochs, with high accuracy as well as Precision and Recall approaching 1.000 for many classes. Although some classification errors persist, they can be minimized through dataset expansion or model parameter refinement. The better validation performance compared to testing also indicates potential overfitting to the training data. Therefore, further analysis of error distribution and optimization of training data can enhance the model's overall performance.

A comparison of the confusion matrix between the validation and testing processes reveals differences in accuracy, classification error rates, and the model's ability to recognize the background, as shown in Figure 9.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802



(a) (b)

Figure 9. Confusion Matrix at 20 Epochs; (a) Validation Process, (b) Testing Process.

During validation, the model achieves an accuracy of 99.84% with an average Precision of 90.7% and Recall of 90.6%. Classification errors occur in class 3 with two errors, and in classes 6 and 9 with one error each. Additionally, the model struggles to recognize the background, as evidenced by Precision and Recall values of 0.000, indicating that all background instances are misclassified as other objects. Meanwhile, in the testing phase, a similar pattern is observed in the confidence curves, where Precision remains stable, but Recall declines at higher confidence thresholds. The classification error rate is higher compared to validation, especially in certain classes. The model also continues to struggle with background recognition, which remains a recurring issue in both evaluation stages.

Overall, the model performs better during validation than testing, suggesting potential overfitting to the training data. Additionally, the faster decline in Recall compared to Precision during testing suggests that the model tends to miss some objects (increased false negatives) when the confidence threshold is raised. Although the model has improved compared to the previous 10 epochs, confusion matrix evaluation indicates that further improvements are needed, such as dataset optimization, data augmentation, or adjusting detection thresholds to reduce classification errors and improve background detection.

A comparison of accuracy, precision, recall, and F1-score for validation and testing over 20 epochs is presented in Table 3.

Table 3. Comparison of Accuracy, Precision, Recall, and F1-score for 20 Epochs

Metric	Validation Process	Testing Process
Accuracy	99,84%	99,91%
Precision	90,70%	91,16%
Recall	90,60%	91,04%
F1-Score	90,70%	90,98%

The final configuration in this study uses 50 epochs to ensure the stability of the YOLOv9 model. Model evaluation was conducted using 163 images with a total of 1,151 instances, achieving a precision of 99.6% and a recall of 99.8%, indicating excellent model performance in detecting objects with minimal error rates. At the individual class level, most classes exhibit near-perfect precision and recall. Classes 0 through 4 have precision close to 100% with perfect recall, indicating that the model makes almost no classification errors in these classes. Some

Telematika: Jurnal Informatika dan Teknologi Informasi
Vol. 22, No. 1, Februari 2025, pp.65-82
ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

classes, such as classes 6 and 7, experience slight reductions in recall or precision, but they remain within a very high range, indicating the model's optimal accuracy and sensitivity.

The confidence curve for the validation process at 50 epochs shows a pattern consistent with previous configurations, as shown in Figure 10.

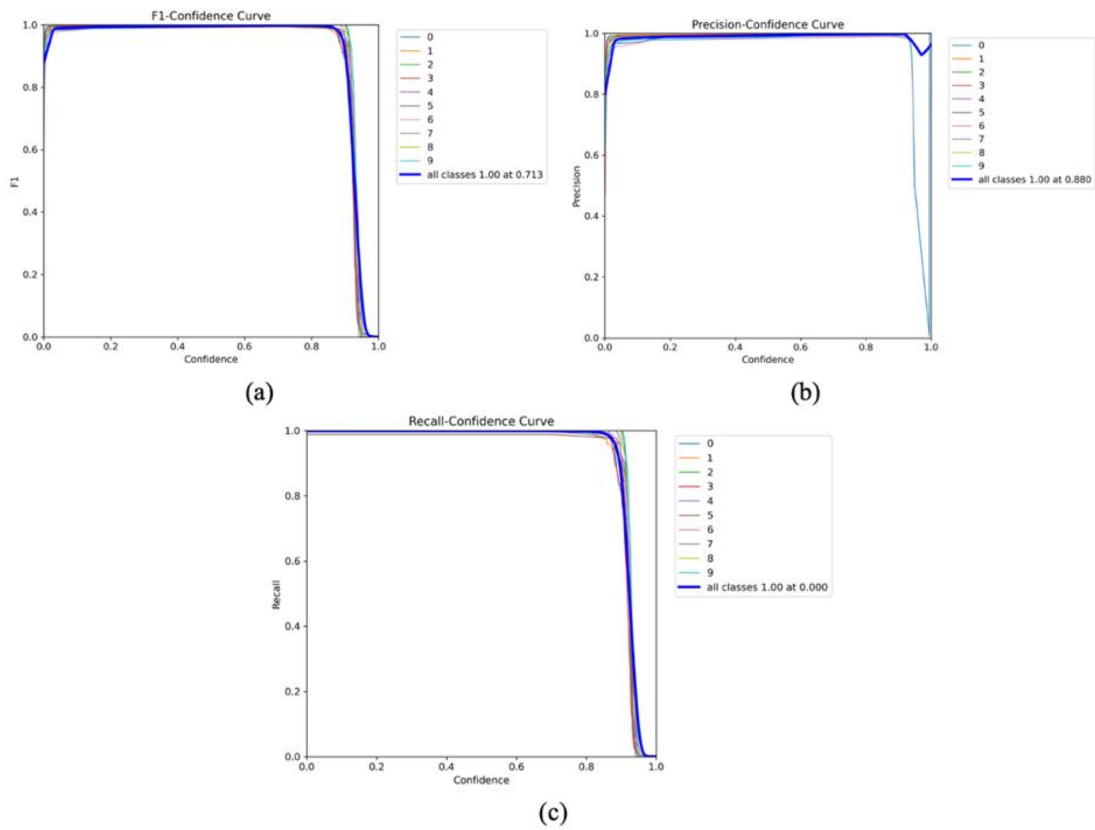
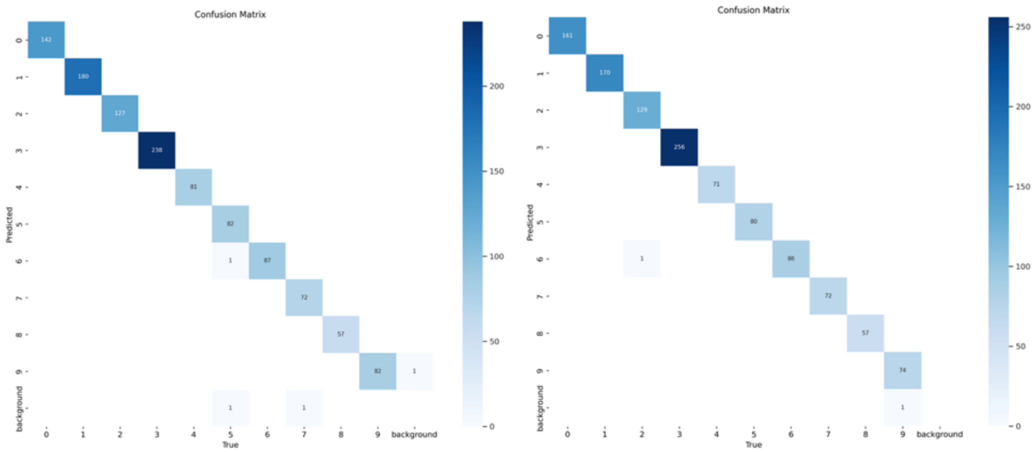


Figure 10. Confidence Curve in the Validation Process at 50 Epochs;
(a) F1-Confidence Curve, (b) Precision-Confidence Curve, and (c) Recall-Confidence Curve

The F1-Confidence Curve graph shows that the F1-score remains high across most confidence ranges before experiencing a sharp decline near a confidence level of 1.0, with an optimal point at a confidence of 0.713. The Precision-Confidence Curve indicates consistently high precision (>0.9) until the confidence level approaches 1.0, with an optimal point at a confidence of 0.880. Meanwhile, the Recall-Confidence Curve also demonstrates good performance, with high recall up to a certain threshold before degrading at high confidence levels. Overall, the model demonstrates a good balance between precision and recall, although there are indications of potential overfitting at high confidence levels.

A comparison of the confusion matrix between validation and testing at 50 epochs, shown in Figure 11, indicates that the YOLOv9 model exhibits very high performance with minimal classification errors in both evaluation stages.



(a) (b)

Figure 11. Confusion Matrix at 50 Epochs; (a) Validation Process, (b) Testing Process

During validation, the model records a total of 1,068 true positives (TP) with only 5 false positives (FP) and 5 false negatives (FN), yielding an overall accuracy of 99.95%. Meanwhile, in the testing process, the model maintains a high accuracy level with a slight increase in TP count, such as in class 3, where 256 instances are correctly classified compared to 238 in validation. However, FP and FN counts remain low, demonstrating the model's consistency in recognizing objects across various classes.

Although overall performance is excellent, there are minor differences in classification error rates between validation and testing. In the validation phase, some classes, such as classes 6 and 9, experience a single misclassified instance, while in the testing phase, class 5 has a similar misclassification error. Additionally, in testing, a case of misclassification occurs where one instance of class 6 is misclassified as class 5, and one instance of class 9 is classified as background. This suggests that although the model has been well-trained, there are still scenarios where it struggles to distinguish between classes with potentially similar features, especially at high confidence levels.

In terms of evaluation metrics such as precision, recall, and F1-score, validation and testing results show nearly identical values, with most classes achieving precision and recall close to 100%. However, classes with classification errors show slight reductions in precision or recall, such as classes 6 and 9, which experience minimal FN and FP in both stages. This stable performance indicates that the model generalizes well to new data, though further optimization can still enhance performance by addressing minor classification errors in testing.

A comparison of accuracy, precision, recall, and F1-score for validation and testing over 50 epochs is presented in Table 4.

Table 4. Comparison of Accuracy, Precision, Recall, and F1-score for 50 Epochs

Metric	Validation Process	Testing Process
Accuracy	99,02%	99,95%
Precision	90,60%	90,70%
Recall	90,60%	90,70%
F1-Score	90,60%	90,70%

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

Based on the evaluation results of the YOLOv9 model in three epoch configurations (10, 20, and 50) for detecting water meter digits, the accuracy comparison of the testing process for each epoch configuration is shown in Table 5.

Table 5. YOLOv9 Model Accuracy Comparison

Configuration	Training Accuracy (%)	Validation Accuracy (%)	Testing Accuracy (%)
10 Epochs	98,56%	99,78%	99,63%
10 Epochs	99,12%	99,84%	99,91%
10 Epochs	99,75%	99,02%	99,95%

Evaluation results indicate that the 20-epoch configuration provides optimal performance with a testing accuracy of 99.91%, an average precision of 91.16%, an average recall of 91.04%, and an average F1-score of 90.98%. Although the 50-epoch configuration achieves the highest accuracy of 99.95%, there is an indication of overfitting, as evidenced by the significant difference between training accuracy (99.75%) and validation accuracy (99.02%). Therefore, the 20-epoch model was selected as it offers the optimal balance between high accuracy, training efficiency, and good generalization to new data.

The confusion matrix shows that the model is capable of recognizing most digit classes effectively. Some classes, such as classes 2 and 4, have perfect precision and recall (1.000), indicating no classification errors. However, there are weaknesses in certain classes, particularly class 9, which has a lower recall (0.910) due to eight false negatives (FN). Although classification errors decrease with more training epochs, some classes, such as classes 5 and 6, still experience slight reductions in precision and recall in the 50-epoch configuration.

The implementation of the YOLOv9 model on a Raspberry Pi 4 Model B microcontroller demonstrates that the model can accurately detect water meter digits in real-world conditions, as shown in Figure 12.

Figure 12. Detection Results Using a Microcontroller

Most digits have a high confidence score above 0.8, but some digits, such as the number 8, have a lower confidence score (0.52), influenced by lighting conditions and image quality. These results indicate that the model is reliable for digit detection with high accuracy. However, challenges such as noise and lighting conditions should be further addressed for real-world applications.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

4. Conclusion

The evaluation of the YOLOv9 model across three epoch configurations (10, 20, and 50) determined that the 20-epoch configuration provided the best balance between accuracy and generalization, achieving an overall accuracy of 99.91%, with an average precision, recall, and F1-score are 91.16%, 91.04% and 90.98%. Although the 50-epoch model reached a slightly higher accuracy of 99.95%, it exhibited overfitting, as indicated by the significant gap between training and validation accuracy. The model also struggled with recognizing the Background class due to the lack of explicit labeling, leading to difficulties in distinguishing digits from environmental noise. Real-world deployment using a Raspberry Pi-based system demonstrated an accuracy of over 95% under optimal lighting conditions but faced challenges in low-light scenarios and with blurred or reflective meter surfaces, affecting detection reliability.

To enhance model performance, future work should focus on improving dataset balance, incorporating Background class labeling, and exploring techniques such as data augmentation, loss function adjustments, and regularization. Additionally, employing alternative models or ensemble learning could further refine detection accuracy. The implementation of this system has proven effective in automating water meter readings, reducing manual input errors, and enabling real-time consumption monitoring. Further refinements can expand its application to broader sectors, such as energy management and IoT-based infrastructure monitoring, ensuring greater efficiency and accuracy in automated numerical recognition systems.

Daftar Pustaka

[1] A. Mittal, “YOLOv9: A Leap in Real-Time Object Detection,” Unite.ai, 5 Maret 2024. [Online]. Available: https://www.unite.ai/yolov9-a-leap-in-real-time-object-detection/. [Diakses 28 November 2024].

[2] A. Jain, D. Somwanshi, K. Joshi dan S. S. Bhatt, “A Review: Data Mining Classification Techniques,” dalam *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, London, United Kingdom, 2022.

[3] S. Sharma dan K. Guleria, “Deep Learning Models for Image Classification: Comparison and Applications,” dalam *2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Greater Noida, India, 2022.

[4] A. Arnita, F. Marpaung, F. Aulia, N. Suryani dan R. C. Nabila, Computer Vision dan Pengolahan Citra Digital, vol. I, A. B. Surya, Penyunt., Surabaya, Jawa Timur: Pustaka Askara, 2022.

[5] F. Indaryanto, A. Nugroho dan A. F. Suni, “Aplikasi Penghitung Jarak dan Jumlah Orang Berbasis YOLO Sebagai Protokol Kesehatan Covid-19,” *Edu Komputika Journal,* vol. 8 (1), no. doi: https://doi.org/10.15294/edukomputika.v8i1.47837, pp. 31-38, 2021.

[6] I. Yuadi, Forensik Digital dan Analisis Citra, Magetan, Jawa Timur: CV. AE Madia Grafika, 2023.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

[7] M. Yaseen, "hat is YOLOv9: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector," *arXiv,* vol. 1, pp. 1-10, 12 September 2024.

[8] Z. Goldfeld dan Y. Polyanskiy, "The Information Bottleneck Problem and its Applications in Machine Learning," *EEE Journal on Selected Areas in Information Theory,* vol. 1 (1), no. doi: 10.1109/JSAIT.2020.2991561, pp. 19-38, Mei 2020.

[9] Y. Wang dan X. Xiang, "GMS-YOLO: an enhanced algorithm for water meter reading recognition in complex environments," *Journal of Real-Time Image Processing (JRTIP),* vol. 21 (173), no. https://doi.org/10.1007/s11554-024-01551-4, 13 September 2024.

[10] W. L. Cheng, "YOLOv9 - A Comprehensive Guide and Custom Dataset Fine-Tuning," 26 Februari 2024. [Online]. Available: https://www.datature.io/blog/yolov9-a-comprehensive-guide-and-custom-dataset-fine-tuning. [Diakses Oktober 2024].

[11] R. F. Putra, I. R. Mukhlis, A. I. Datya, S. J. Pipin, F. Reba, M. Al-Husaini, S. A. Mandowen, N. N. L. E. Zain dan L. Judijanto, Algoritma Pembelajaran Mesin: Dasar, Teknik, dan Aplikasi, PT. Sonpedia Publishing Indonesia, 2024.

[12] A. Amwin, "Deteksi Dan Klasifikasi Kendaraan Berbasis Algoritma You Only Look Once (Yolo)," Universitas Islam Indonesia, Yogyakarta, 2021.

[13] R. Rina, "Memahami Confusion Matrix: Accuracy, Precision, Recall, Specificity, dan F1-Score untuk Evaluasi Model Klasifikasi," 12 Juni 2023. [Online]. Available: https://esairina.medium.com/memahami-confusion-matrix-accuracy-precision-recall-specificity-dan-f1-score-610d4f0db7cf. [Diakses 20 November 2024].

[14] M. I. Ardiansyah, "Implementasi You Only Look Once v5 (YOLOv5) untuk Identifikasi Genus Anggrek di Indonesia Berbasis Mobile," Universitas Lampung, Bandar Lampung, 2023.

[15] Y. A. Singgalen, "Implementasi Metode CRISP-DM dalam Analisis Model Pendukung Keputusan Simple Additive Weighting dan Pengembangan Basis Data Riwayat Pembelian Layanan Akomodasi Hotel," *Jurnal Sistem Komputer dan Informatika (JSON),* vol. 5, no. 2, p. 308−317, Desember 2023.

[16] S. Navisa, L. Hakim dan A. Nabilah, "Komparasi Algoritma Klasifikasi Genre Musik pada Spotify Menggunakan CRISP-DM," *Jurnal Sistem Cerdas,* vol. 04, no. 02, pp. 114-125, 2021.

[17] R. Laroca, V. Barroso, M. A. Diniz, G. R. Goncalves, W. R. Schwartz dan D. Menotti, "Convolutional Neural Networks for Automatic Meter Reading," *Journal of Electronic Imaging,* vol. 28 (1), no. Doi: 10.1117/1.JEI.28.1.013023, pp. 1-38, 2019.

[18] G. Salomon, R. Laroca dan D. Menotti, "Deep Learning for Image-based Automatic Dial Meter Reading: Dataset and Baselines," *Arxiv,* vol. 2, no. doi: https://doi.org/10.48550/arXiv.2005.03106, pp. 1-8, 8 Mei 2020.

ISSN: 1829-667X / E-ISSN: 2460-9021
DOI:10.31515/telematika.v22i1.14802

[19] Q. Hong, Y. Ding, J. Lin, M. Wang, Q. Wei, X. Wang dan M. Zeng, “Image-Based Automatic Watermeter Reading under Challenging Environments,” *Sensors,* vol. 21 (2), no. Doi: https://doi.org/10.3390/s21020434, pp. 1-18, 9 Januari 2021.

[20] B. Widodo, H. Armanto dan E. Setyati, “Deteksi Pemakaian Helm Proyek Dengan Metode Convolutional Neural Network,” *INSYST: Journal of Intelligent System and Computation* , vol. 3 (1), p. 23–29, April 2021.

[21] R. I. Ilham dan F. Utaminingrum, “Deteksi Helm untuk Keamanan Pengendara Sepeda Motor dengan Metode CNN (Convolutional Neural Network) menggunakan Raspberry Pi,” *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,* vol. 5 (11), no. e-ISSN: 2548-964X, pp. 4734-4739, November 2021.

[22] Roboflow, “Roboflow Annotate: The Fastest Way to Label Computer Vision Datasets,” Roboflow, Inc., [Online]. Available: https://roboflow.com/annotate. [Diakses Desember 2024].

[23] T. Kosasi, Z. A. L. Sihombing dan A. M. Husein, “YOLO-Based Vehicle Detection: Literature Review,” *Journal of Computer Networks, Architecture and High Performance Computing,* vol. 6, no. 3, pp. 1384-1389, doi: https://doi.org/10.47709/cnapc.v6i3.4377, Juli 2024.