

Image Enhancement Technique Utilizing YOLO Model for Automatic Number Plate Recognition



Christine Dewi¹, Stephen Abednego Philemon¹, Guowei Dai², Abbott Po Shun Chen^{3*}

¹ Department of Information Technology, Satya Wacana Christian University, Salatiga 50711, Indonesia

² College of Computer Science, Sichuan University, Chengdu 610065, China

³ Department of Marketing and Logistics Management, Chaoyang University of Technology, Taichung 413310, Taiwan

Corresponding Author Email: chprosen@gm.cyut.edu.tw

Copyright: ©2025 The authors. This article is published by IETA and is licensed under the CC BY 4.0 license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.18280/ijtdi.090106>

ABSTRACT

Received: 18 November 2024

Revised: 11 February 2025

Accepted: 24 February 2025

Available online: 31 March 2025

Keywords:

Automatic Number Plate Recognition (ANPR), YOLOv9, EasyOCR, image enhancement, real ESRGAN, CLAHE

The significant increase in Indonesian vehicle numbers has highlighted the importance of a robust ANPR system. Over the past few years, the number has increased by up to 4% each year and is expected to continue rising as long as economic growth continues. This study utilized the YOLOv9 model and EasyOCR along with image enhancement as a pipeline for the license plate recognition process. YOLOv9 was chosen for object detection because its architecture offers good stability in terms of performance and efficiency, even outperforming newer models with a 99.3% mAP@50 in the YOLOv9s variant. EasyOCR was used to recognize and extract characters from license plates. To enhance image quality, Real-ESRGAN upscales image resolution using a GAN architecture and addresses character blurring caused by low-resolution images. Additionally, CLAHE further enhances the clarity of low-contrast characters by employing a histogram to redistribute image intensities. The proposed method achieved an accuracy of 84.36% when tested on 100 image samples in real-world situations, indicating fairly good performance despite challenges like blurring and low contrast. The results highlight the potential of ANPR solutions in addressing the common challenges of real-time license plate recognition, contributing to more efficient traffic surveillance and enforcement systems, especially in Indonesia.

1. INTRODUCTION

Indonesia's vehicle count has been rising steadily, driven by the country's economic growth and growing population needs. As one of the world's most populous nations, Indonesia faces substantial challenges in managing traffic. Data from the Central Bureau of Statistics highlights this upward trend, with vehicle numbers climbing consistently year after year. In 2022, there were 125,305,332 motorcycles in Indonesia, with an annual growth rate of 4.11%, while the number of cars increased by 3.73% to 17,168,862 units [1].

A vehicle license plate is an important thing for every vehicle user who drives on public highways must have. Each vehicle has a unique license plate number as an identifier for traffic surveillance and law enforcement purposes. The high accidents, vehicle theft, and traffic violations in Indonesia makes it difficult for law enforcement to conduct manual surveillance [2, 3]. This emphasizes how important automated traffic surveillance systems are in ensuring more effective law enforcement. Nowadays, Electronic Traffic Law Enforcement (ETLE) is an innovative automated traffic surveillance technology that can reduce the manual law enforcement. This technology was first introduced to Indonesia in 2018 and can automatically recognize vehicle license plates through

Automatic Number Plate Recognition (ANPR), which can improve the efficiency of traffic surveillance [4, 5].

Automating number plate recognition systems could be possible with advances in artificial intelligence and computer vision. An object detection model called YOLO (You Only Look Once) can be used to detect license plates on Indonesian vehicles. Furthermore, optical character recognition (OCR) can be utilized to recognize and extract characters from license plates. However, real-world challenges such as Indonesia's diverse weather conditions, different license plate angles, and the wide variety of Indonesian license plate styles can lead to recognition issues with this technology.

One approach to addressing these problems is to apply several image-enhancement techniques [6]. Real-ESRGAN and CLAHE are two techniques for improving visual image quality. Real-ESRGAN elevates image resolution, and CLAHE adjusts the image contrast, helping to reduce character recognition errors and making OCR more accurate. This study investigates the implementation of YOLOv9 to detect Indonesian vehicle license plates, as well as the application of image quality enhancement methods, and to evaluate how this approach may increase the accuracy of EasyOCR, especially in real-world conditions.

2. RELATED WORK

2.1 YOLO algorithm

YOLO’s primary objective is to instantly locate and identify objects in real time. This object detection model uses a single neural network, which makes it fast and effective. In Figure 1, YOLO divides the image into $S \times S$ grids [7]. Eq. (1) is used to calculate a confidence score $\text{Confidence}(C)$, which predicts the presence of objects in each grid cell [8]. Based on the probability values $P(\text{Object})$ from the class probability map, cells containing object parts are combined into a unified bounding box per detected object. The YOLO approach uses non-maximum suppression, eliminating duplicates while minimizing bounding box mistakes [9].

$$\text{Confidence}(C) = P(\text{Object}) \times \text{IOU}_{\text{pred}}^{\text{truth}} \quad (1)$$

The intersection over Union $\text{IOU}_{\text{pred}}^{\text{truth}}$ metric measures how closely the predicted bounding box matches the actual location of an object in the image. This metric divides the overlapping area between the predicted and actual boxes ($A \cap B$) by their total area ($A \cup B$). Figure 2 shows the varying IoU values based on the extent of overlap between both boxes, a higher IoU value indicates that the prediction better matches the actual location of the object [9]. For example, a low IoU value of 0.33 indicates minimal overlap between the predicted and actual boxes, whereas a nearly perfect overlap is represented by a high IoU of 0.95.

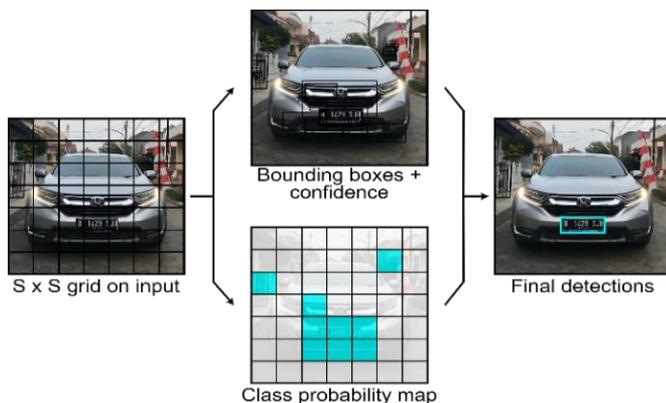


Figure 1. YOLO framework

The implementation of YOLO for license plate detection in ANPR systems has shown excellent performance in recent years. YOLOv8 was successfully utilized by Sarhan et al. [10] to detect Egyptian license plates in real time. Furthermore, Tran-Anh et al. [11] confirmed that YOLOv8 is effective in detecting license plates from a variety of camera angles. Compared to earlier YOLO models like YOLOv8, the deep learning-based object detection methods in YOLOv9 or YOLOv10 have been shown to be more accurate and effective in real-time detection of objects, making them expected to play a crucial role in the real-time detection [12, 13].

YOLOv9 advances its architecture with two key innovations that enhance performance and efficiency. First, Programmable Gradient Information (PGI) addresses the issue of deep supervision. This issue often creates information bottlenecks that are incompatible with lightweight neural network architectures. The bottlenecks can lead to data loss

during transformations in deep neural networks, resulting in unreliable gradients that has a negative impact on the object detection performance. PGI enables the deep neural network to generate more reliable gradients and is appropriate for implementation in lightweight architectures. Second, Generalized Efficient Layer Aggregation Network (GELAN) architecture combines CPSNet with ELAN to improve the stability and performance of the model [14, 15].

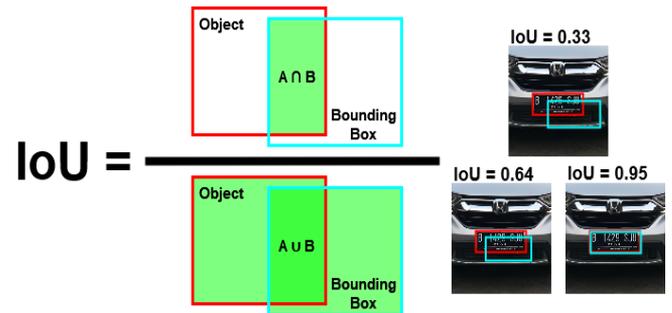


Figure 2. How to calculate IoU

2.2 Optical character recognition

Optical character recognition (OCR) is a technology that detects and transforms text from images into readable digital text. OCR uses advanced artificial intelligence and image processing technology to extract characters or text from images. EasyOCR was produced by Jaided AI (2020), which is a popular OCR tool [16]. EasyOCR is well-regarded for being user-friendly and delivering impressive results in extracting text from images. It combines the CRAFT method for detecting text and the CRNN for recognizing individual characters [17, 18]. The CRAFT method uses a Regional Proposal Network (RPN) to detect text regions and a Character Proposal Network (CPN) to improve accuracy by predicting bounding boxes around the text. ResNet and LSTM networks are used to extract features from text, allowing EasyOCR to recognize characters in images [10]. The findings of Vedhaviyassh et al. [19] show that EasyOCR is highly effective for real-time vehicle license plate detection, outperforming Google's TesseractOCR.

2.3 Image enhancement method

Image enhancement is a method for improving and adjusting the visual quality of images. The quality of input images for OCR in real-time Indonesian license plate detection varies according to the environment. Blurry images, high noise levels, and uneven contrast are the leading factors of inaccurate OCR predictions. Researches by Kabiraj [20] and Wang et al. [21] have demonstrated that enhancing image resolution and contrast yields impressive results in improving the accuracy of OCR.

Enhanced Super-Resolution Generative Adversarial Networks (ESRGAN) builds upon SRGAN, which utilizes a GAN architecture where the generator and discriminator work concurrently to significantly enhance image resolution. The discriminator evaluates the high-resolution images produced by the generator. The ESRGAN generator comprises multiple Residual-in-Residual Dense Blocks (RRDBs) that are built up from Dense Block layers. As shown in Figure 3, Dense Blocks use convolutional layers and Leaky ReLU activations to enhance the flow of information and gradients, help prevent

vanishing gradient issues and allow the construction of deeper networks [22]. The generator then upscales the image using a series of upsampling layers, producing a high-resolution

output image. This architecture transforms low-resolution images into high-resolution ones, resulting in smooth textures and detailed image quality.

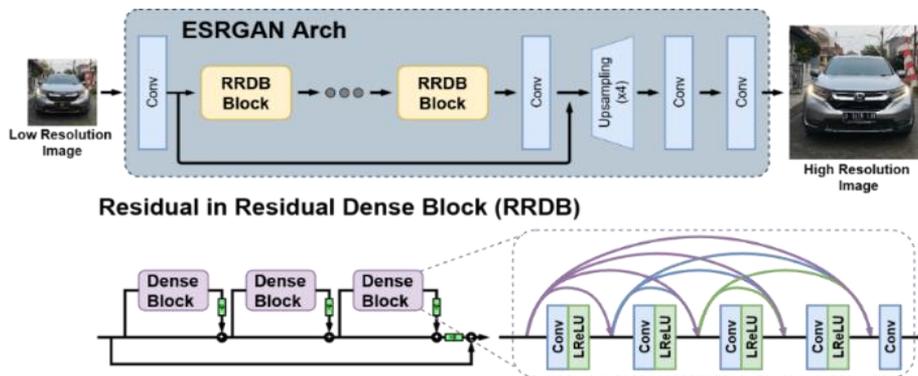


Figure 3. ESRGAN generator

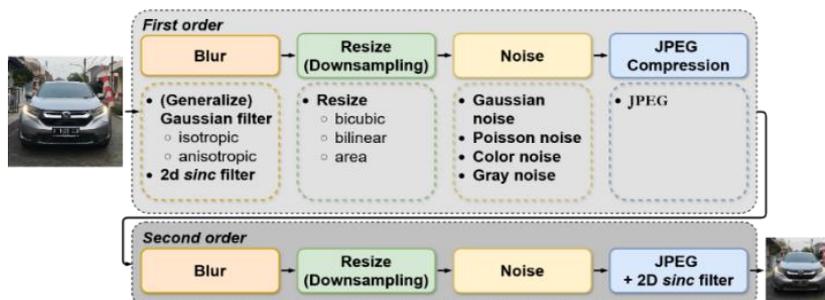


Figure 4. Degradation model in Real-ESRGAN

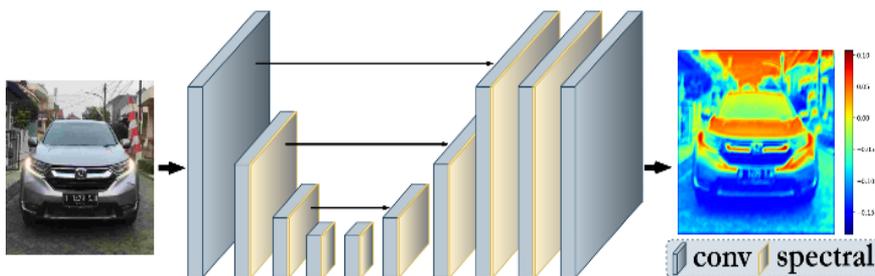


Figure 5. Real-ESRGAN discriminator with U-Net and spectral normalization

Real-ESRGAN is a version of ESRGAN that keeps the same generator architecture but is optimized to enhance image resolution in real-world conditions. This version extends the original $\times 4$ ESRGAN architecture to scale factors of $\times 2$ and $\times 1$, adapting more effectively to handle a wider range of image resolutions. Consequently, the architecture can become a heavy network, and pixel unshuffle must be employed at the new scale factors to reduce spatial dimensions and expand the number of channels. Furthermore, Real-ESRGAN applies an advanced degradation model to convert high-resolution images into synthetic data (see Figure 4) [23]. The degradation model allows the training phase to include more diverse data that reflects real-world conditions, in order to address its challenges. The complexity of its degradation model causes the discriminator architecture used in ESRGAN to be ineffective, so Real-ESRGAN employs a U-Net structure with spectral normalization to effectively manage this problem. The U-Net structure includes a skip connection mechanism, allowing it to generate reliability scores for each image pixel and provide them to the generator [24]. The spectral normalization regularization applied to the U-Net structure

improves stability during the training process. This regularization method keeps the discriminator from becoming too sensitive to variations in the input based on real-world noise and artifacts introduced during degradation [25]. As shown in Figure 5, spectral normalization and U-Net’s hierarchical feature extraction allow Real-ESRGAN to effectively handle different distortions while ensuring a stable training process [23]. The mechanism generates high-resolution images with significant improvement, even in challenging real-world conditions.

Contrast Limited Adaptive Histogram Equalization (CLAHE) enhances image contrast, particularly in low-contrast images, by employing histogram equalization [26]. The histogram maps the distribution of pixel intensities in the image. As suggested by its name, CLAHE uses a limiting mechanism called the clip limit to restrict the maximum pixel intensity. If any intensity value exceeds the clip limit, it is equalized to remain within the limit. Dewi et al. [27] demonstrated the good performance of CLAHE in enhancing contrast in low-contrast images compared to other contrast enhancement techniques. Similarly, Chang et al. [28] found

comparable results, even though CLAHE does not always outperform other techniques in all situations.

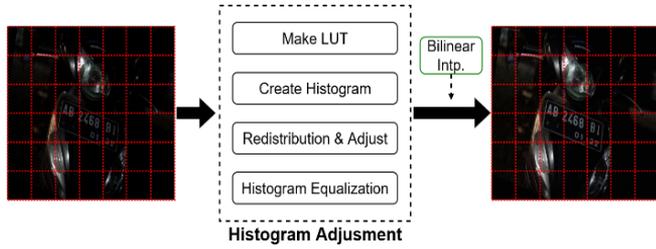


Figure 6. CLAHE framework

Figure 6 shows the steps for doing histogram adjustment with CLAHE, which are as follows [28, 29]: (1) Split the input image into small blocks; (2) Create a histogram for each block and limit the maximum pixel intensity using a clip limit; (3) Redistribute the pixel values that exceed the limit evenly to the lower part of the histogram; (4) Equalize the histogram to adjust the intensity distribution of grayscale levels within each block. Each pixel intensity within a block has a corresponding probability, which can be determined using the Probability Density Function (PDF) as follows:

$$PDF(j) = \frac{n_j}{n}, j=0, 1, 2, \dots, L-1 \quad (2)$$

where, j denotes the intensity level in a grayscale image, where $L-1$ (or 255) is the maximum value. n_j represents the number of pixels at intensity level j , and n refers to the total number of pixels within the block. Based on $PDF(j)$, CLAHE uses histogram equalization to adjust the pixel intensities within each block. This process redistributes the intensity values, making sure that every image block has a more consistent intensity distribution, thereby enhancing the overall contrast. The distribution is then calculated using the Cumulative Distribution Function (CDF) as follows:

$$CDF(k) = \sum_{j=0}^k PDF(j) \quad (3)$$

Then, the gray levels of the image blocks can be remapped using the following equation:

$$s_k = T(k) = (L-1) \times CDF(k) \quad (4)$$

s_k is the output of the remapping function $T(k)$ that determines the new pixel intensity value based on $CDF(k)$, where, k is the pixel value in the gray intensity level.

Bilinear interpolation in CLAHE prevents blocking artifacts introduced during the remapping process. This method interpolates new pixel intensity values based on the grayscale values of surrounding pixels. If the grayscale value at sample point P is k , and the interpolated pixel value is k' , the sample point is surrounded by the points P_a , P_b , P_c , and P_d . The corresponding remapping of the grayscale levels from k is denoted as $s_{P_a}(k)$, $s_{P_b}(k)$, $s_{P_c}(k)$, and $s_{P_d}(k)$. The interpolated pixel value at the grayscale level k' can be calculated from the following equation:

$$k' = (1-\beta)((1-\alpha) \cdot s_{P_a}(k) + \alpha \cdot s_{P_b}(k)) + \beta((1-\alpha) \cdot s_{P_c}(k) + \alpha \cdot s_{P_d}(k)) \quad (5)$$

where, α and β are the normalized distances with respect to the point P_a .

3. METHODOLOGY

In this study, we utilize the YOLOv9 model to detect Indonesian vehicle license plates and employ EasyOCR to extract the characters from them. Both YOLOv9 and EasyOCR are known for their good performance within the scope of ANPR technology. YOLOv9 is a state-of-the-art model for real-time object detection, while EasyOCR offers outstanding accuracy and ease of use in recognizing alphanumeric characters. The performance of EasyOCR is further improved by the application of image enhancement techniques, including Real-ESRGAN and CLAHE. This combination makes the system both reliable and efficient. The overall workflow of this research is illustrated in Figure 7.

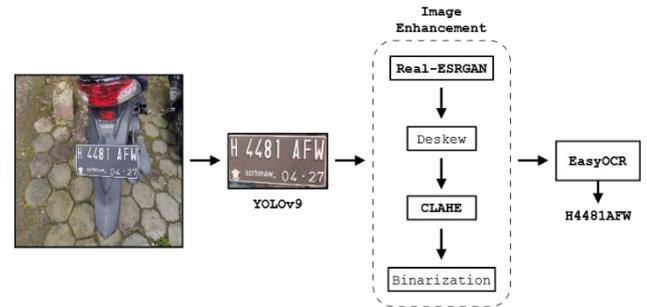


Figure 7. Research workflow

3.1 Object detection with YOLOv9

YOLOv9 is ideal for license plate real-time detection with high accuracy and efficiency in detecting small objects like license plates, which makes it well-suited for this task. The cropping process begins when YOLOv9 detects a license plate, extracting the bounding box area for further processing to obtain a Region of Interest (ROI) [30]. This step isolates the license plate from the rest of the image, not only to take the relevant part of the image but also to reduce the computational load. Consequently, it will improve OCR's ability in detecting and recognizing characters on the license plates. The YOLOv9 model is trained and evaluated to ensure that it operates with high detection accuracy.

The training process is intended to teach the YOLOv9 models about the characteristics of vehicle license plate objects, allowing them to improve their performance in detecting these objects. The training process begins with data preprocessing that resizes the images to a consistent size for maintaining consistency across the dataset. The training data also applies augmentation techniques, including random flipping, brightness adjustments, and scaling, to further improve the model's robustness and ability to handle varying conditions. We select the training hyperparameters based on the work of Singh et al. [31] and Amin et al. [32], whose studies highlighted an effective balance between accuracy and efficiency for license plate detection (see Table 1). Their findings serve as a foundation for optimizing our model's performance. The learning rate is selected in conjunction with the AdamW optimizer to ensure training stability while enabling efficient updates to the model's weights. The training batch size can be adjusted based on both GPU memory

utilization and the benefits of gradient updates. The chosen epoch and weight decay help the training process to prevent overfitting. Image size 640 is a standardized size to balance the ability to capture image information while maintaining computational efficiency. The chosen momentum is a standardized value that also stabilizes the training process and ensures the model achieves a minimum loss faster. These hyperparameters are adjusted to help the model adapt more effectively to real-world conditions based on the characteristics of the license plate dataset. Figure 8 is an example of the ROI obtained with YOLOv9.



Figure 8. An example of the ROI obtained with YOLOv9

Table 1. Hyperparameters for training the model

Hyperparameters	Value
Learning rate	0.01
Batch size	16
Epochs	50
Image size	640
Optimizer	AdamW
Momentum	0.937
Weight decay	0.005

The YOLOv9 models are evaluated to assess their performance in detecting license plates. The evaluation process involves a detailed examination of the models' ability to understand how well they perform. The evaluation utilizes standard metrics that provide insight, which are defined in Eqs. (6)-(8):

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

where, TP (True Positive) refers to the number of objects that are correctly detected by the model, and FP (False Positive) indicates the number of instances where an object is incorrectly detected as a license plate. Precision measures the proportion of correctly detected license plates among all instances that were labeled as license plates by the model. A higher precision value indicates that the detected license plates are truly correct.

$$\text{Recall} = \frac{TP}{TP + FN} \quad (7)$$

where, FN (False Negative) refers to the number of objects that are present in the image but are missed or not detected by the model. Recall measures how effectively the model can detect every important object in the image, including those that might be challenging to detect. A high recall value indicates that the model successfully detects a large proportion of the true objects in the image.

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i \quad (8)$$

where, N represents the overall count of object categories, while AP_i denotes the Average Precision for each specific category i . AP is a key metric that calculates the precision-recall curve for a category, then summarizes it into a single value. mAP calculates the mean AP across all categories at a given IoU threshold (mAP@50 refers to a minimum IoU of 50%).

3.2 Image enhancement with Real-ESRGAN and CLAHE

Image enhancement will improve image quality after the ROI is obtained. This is important because the cropping process can reduce the image's visual quality, making it harder for OCR to accurately detect and recognize characters on the license plate. Real-ESRGAN and CLAHE are the primary techniques applied in this study to enhance the clarity and detail of license plate images. Real-ESRGAN enhances image resolution and sharpens the characters on the license plate, while CLAHE adjusts contrast distribution in the images to achieve more even contrast. Additional techniques, including deskewing and binarization, are also used to support this method.

3.2.1 Real-ESRGAN

Real-ESRGAN is the initial phase in our image enhancement process. In this work, we utilize a scaling factor of $\times 4$ to upscale the resolution of license plate images. The upsampled resolution enhances the clarity of characters that were previously blurred in low-resolution images. Real-ESRGAN was chosen because it can improve the visual quality of real-world images without adding noise or artifacts, allowing even the most challenging license plate images to be processed correctly. This means that the technique can enhance image visuals for better recognition of small and intricate details that are essential for accurate OCR processing, while preserving important features.

3.2.2 Deskewing

Before deskewing, the enhanced images will first be resized again to 920×430 pixels using bilinear interpolation to maintain uniformity across all input images. Furthermore, the deskewing process begins by determining the image's tilt angle. The determined tilt angle is then applied to correct the skew, ensuring that it does not interfere with character detection and recognition. We use the Hough Line Transform, a method for identifying straight lines in images, to calculate the angle. This method calculates the distance between the point $(0, 0)$ and the straight line, known as the normal line, using the following equation [33]:

$$x \cos \theta + y \sin \theta = \rho \quad (9)$$

where, x and y represent the Cartesian coordinates of the boundary point, and θ denotes the angle between the normal line and the x -axis. The Hough Line Transform computes a two-dimensional parameter space (θ, ρ) to find the dominant θ value. The image's tilt angle is determined by identifying the most frequent θ value from the mapped coordinates in the parameter space. This angle is then used to rotate the skewed license plate image relative to a defined central position using

the two-dimensional rotation matrix [34]. The image's point (x, y) is rotated by an angle θ using Eq. (10), creating a new point (x', y') .

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad (10)$$

3.2.3 CLAHE

The local contrast of the vehicle license plate image is adjusted using Contrast Limited Adaptive Histogram Equalization (CLAHE), which divides the image into small 8×8 blocks. In this study, histogram equalization is applied to these blocks with an adaptive clip limit, calculated by the following equation:

$$\text{ClipLimit}(i) = \frac{\alpha}{\sigma_i + c} \quad (11)$$

where, α is the contrast control parameter, σ_i is the standard deviation of image i , and c is a small constant to prevent division by zero. This adaptive method ensures more precise contrast adjustment according to the specific conditions of the image.

The value of α is set to 50 to adjust the image contrast. The CLAHE process is performed by separating the RGB image into three individual color channels: Red, Green, and Blue. Each color channel is then processed independently using CLAHE, aiming to adjust the local contrast of each channel. After that, the three-color channels are combined back together to form an RGB image that has been adjusted for contrast as the output.

3.2.4 Binarization

Binarization simplifies the process of OCR character extraction from vehicle license plate images. Otsu's binarization method automatically determines the optimal threshold value based on the gray level histogram by dividing it into two distinct classes: background C_0 and object C_1 . Let the probability distribution value at gray level i in this method be denoted as p_i . The probabilities of class occurrence and the class mean levels are calculated as follows [35]:

$$\omega_{C_0} = \sum_{i=0}^k p_i, \quad \omega_{C_1} = \sum_{i=k+1}^{L-1} p_i \quad (12)$$

and

$$\mu_{C_0} = \frac{\sum_{i=0}^k ip_i}{\omega_0}, \quad \mu_{C_1} = \frac{\sum_{i=k+1}^{L-1} ip_i}{\omega_1} \quad (13)$$

where, $\omega_{C_0} = \omega(k)$ is the cumulative distribution value at threshold k . The inter-class variance is then calculated as follows:

$$\sigma_B^2(k) = \omega_0(k)\omega_1(k)(\mu_1(k) - \mu_0(k))^2 \quad (14)$$

The optimal threshold value k^* is obtained by maximizing $\sigma_B^2(k)$, which is calculated using the following equation:

$$\sigma_B^2(k^*) = \max_{0 \leq k < L-1} \sigma_B^2(k) \quad (15)$$

Otsu's binarization technique segments the image into foreground (character areas on the license plate) and background by evaluating the pixel values and the threshold k^* . The input image $T(X, Y)$ is segmented based on the threshold k^* in the following manner [36]:

$$G(X, Y) = \begin{cases} 1 & \text{if } T(X, Y) \geq k^* \\ 0 & \text{if } T(X, Y) < k^* \end{cases} \quad (16)$$

where, $G(X, Y)$ is a function that groups the pixel intensities from $T(X, Y)$ into two feature vectors. This function has the criterion that if the intensity value of the pixel $T(X, Y)$ in gray level is greater than or equal to the threshold k^* , then that pixel is considered an object. In contrast, if the intensity value of the pixel is less than the threshold k^* , then the pixel is considered part of the background.

3.3 Testing and evaluating with EasyOCR

After all previous processes are completed, EasyOCR is the final step in detecting and recognizing the characters on the license plate, converting them into digital text. The performance of EasyOCR is then evaluated using an accuracy metric to see the effect of each image enhancement technique applied to the license plate images. This evaluation examines how different image enhancement techniques, such as resolution upscaling and contrast adjustment can affect the accuracy of the character recognition process. The accuracy metric is calculated by comparing the predicted characters with the ground truth from the original image using Eq. (17).

$$\text{Accuracy} = \frac{n_{\text{truth}} - \text{LD}}{n_{\text{truth}}} \times 100 \quad (17)$$

where, n_{truth} is the number of characters in the ground truth, and Levenshtein Distance LD is the minimum number of character changes required to transform the predicted text into the actual text. This includes additions, deletions, or substitutions of characters, allowing for an evaluation of the prediction accuracy [37].



Figure 9. Examples of Indonesia LPR Dataset

3.4 Dataset

The Indonesia LPR Dataset from Roboflow was created primarily for vehicle license plate recognition [38]. This dataset comprises a variety of images of Indonesian vehicles, including their license plates (see Figure 9). The images provided include variations in contrast conditions, viewpoints, and vehicle plate types. Furthermore, the images are taken in

a variety of settings such as roads, parking lots, and public areas. To match with the research purposes, we limit our emphasis to the license plate class within this dataset. It contains a total of 7,913 images, divided into three categories: 6,203 for training, 846 for validation, and 864 for testing.

4. RESULTS AND DISCUSSIONS

This section presents the results of implementing Indonesian license plate recognition in real time, utilizing YOLOv9 for license plate detection, and applying image enhancement methods to improve character recognition accuracy with EasyOCR. The evaluation of each method provides an overview of the performance of license plate recognition, especially in conditions that align with real-world scenarios.

The entire training process utilized P100 GPU computation available on the Kaggle platform. Three variations of the YOLOv9 model were used for training: YOLOv9s, YOLOv9m, and YOLOv9c, along with more recent object detection models from YOLOv10 and YOLO11, including YOLOv10s, YOLOv10m, YOLO11s, and YOLO11m. Figure 10 shows how effectively these models performed when the hyperparameters described in Table 1 were applied. The steadily decreasing loss values indicate the model got more accurate at detecting license plates. Metrics detailed in Section 3.1 were used to evaluate each model, and the results show that YOLOv9s achieved the highest mAP@50 values at 99.3% (see Table 2). This model has a better result not only in mAP@50, but also in precision and mAP50-95. Apart from that, the comparative analysis between YOLOv9, YOLOv10, and YOLO11 explains that the ‘small’ models have more balanced results in precision and mAP@50-95 compared to larger models, such as the ‘medium’ models. These results suggest how well the smaller models perform under different conditions that might appear in ANPR tasks, especially in Indonesian areas. Smaller models also offer more

computational efficiency. It is really ideal for real-time tasks that require both fast processing and high accuracy.

The EasyOCR evaluation was conducted by taking 100 sample images of Indonesian vehicles. YOLOv9s was applied to each sample image to detect license plate objects and obtain the ROI. Next, the visual quality of the ROI image is enhanced with the image enhancement methods described in Section 3.2, as shown in Figure 11. First, Real-ESRGAN mitigates the impact of low-resolution surveillance cameras and motion blur, which is often caused by fast-moving vehicles. The blurring effect on characters can be significantly reduced for better readability by enhancing image resolution. Secondly, CLAHE specifically handles underexposed license plate objects at night or in other low-contrast situations. Furthermore, deskewing and binarization are used as the support techniques. Deskewing handles skewed license plate that might happen while implementing real-time recognition. This technique can be considered an important step because OCR systems generally read characters in a straight line. Finally, binarization is used to separate characters from the background part. In summary, these enhancement techniques can make characters on the license plates more legible and improve the overall accuracy of the following character recognition process.

The evaluation results in Table 3 clearly demonstrate an improvement in accuracy as each enhancement method is applied. According to the evaluation results of 100 tested samples, EasyOCR can achieve the highest average accuracy of 84.36%, while deskewing and binarization alone only achieve an average accuracy of 49.76%. Real-ESRGAN greatly increased the average accuracy of EasyOCR by enhancing the details of license plates in low-resolution images, indicating this technique is effectively refining characters in the license plate area in real-world scenarios, while some low-contrast image samples were effectively adjusted using CLAHE technique. This combination can significantly improve the accuracy of character detection and recognition by EasyOCR (see Figure 12).

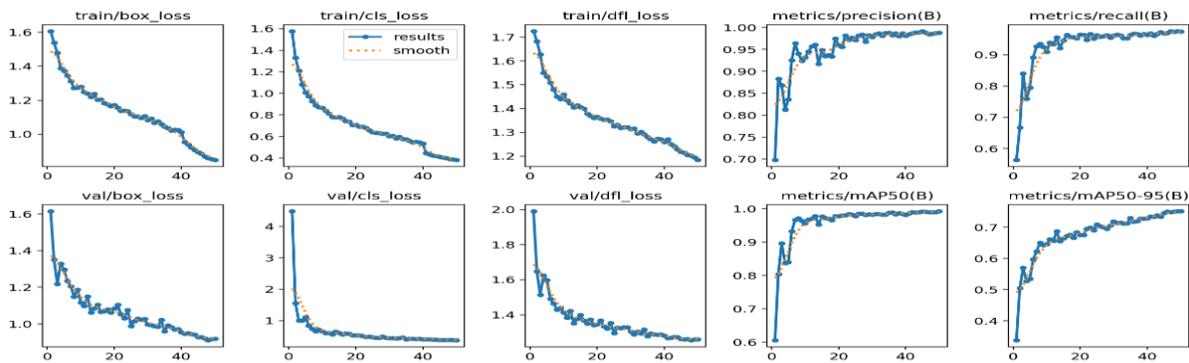


Figure 10. Training results for YOLOv9s

Table 2. Model evaluation results

Model	P	R	mAP@50	mAP@50-95
YOLOv9s	0.988	0.974	0.993	0.753
YOLOv9m	0.984	0.976	0.988	0.744
YOLOv9c	0.985	0.965	0.988	0.746
YOLOv10s	0.985	0.952	0.987	0.744
YOLOv10m	0.973	0.956	0.988	0.74
YOLO11s	0.982	0.967	0.987	0.741
YOLO11m	0.974	0.976	0.989	0.735

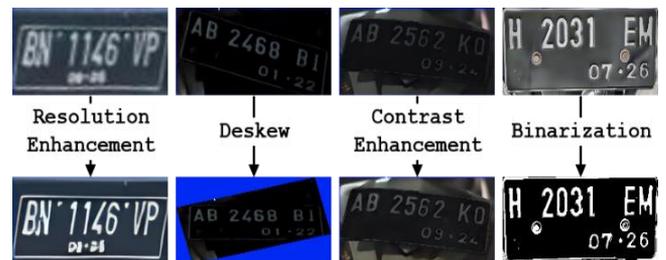


Figure 11. Examples of image enhancement methods



Figure 12. EasyOCR test results with the proposed method

Table 3. Evaluation of EasyOCR with each image enhancement method

Methods	Average Accuracy
Only Deskewing and Binarization	49.76%
Real-ESRGAN	78.67%
Real-ESRGAN+CLAHE	84.36%

Although EasyOCR can achieve relatively high accuracy with the proposed method, there are some limitations that need to be discussed. The proposed method focuses on detecting and recognizing Indonesian vehicle license plates under Indonesian regional conditions. The further adjustment is needed to perform optimally in regions with significantly different license plate designs. However, it remains an unresolved challenge where the vehicle license plates are physically damaged, including cracks, scratches, or degraded. The current method relies solely on improving visual clarity, and cannot address physical damage factors. Apart from that, Table 4 highlights the limitation of YOLOv9 without image enhancement techniques, which struggles to detect license plate objects in a low-contrast condition. This implies that the application of several image enhancement techniques is also important in the detection process. Table 4 also highlights the limitation of applied EasyOCR, which frequently misrecognizes certain characters on Indonesian vehicle license plates. For example, characters such as 'Q' are frequently mistaken for 'O', 'O' for '0', and 'G' for '6'. These errors are generally caused by the visual similarities between the characters. Additional training to fine-tune EasyOCR using a dataset specifically composed of Indonesian license plate characters could improve its ability to distinguish them.

Table 4. Detailed test results of EasyOCR with the proposed method

License Plate Images (ROI)	Image Size	Detected by YOLOv9	Ground Truth	EasyOCR Prediction	Accuracy
	71×45	Yes	BK1091ACY	BK1091ACY	100.0%
	83×70	Yes	G9506PA	69506PA	85.71%
	69×35	Yes	R1599QR	R1599OR	85.71%
	658×356	Yes	AB3874OT	AB38720T	75.0%
	53×39	Yes	AB2833SZ	B2833	62.5%
	53×39	Yes	B888SML	6886SHL	57.14%
	640×640	No	BH3485CI	-	0%

5. CONCLUSIONS

The rising number of vehicles in Indonesia has emphasized the need for automated traffic surveillance systems to support law enforcement. Various methods from previous studies have been proposed to improve their performance. As one of them, the integration of Real-ESRGAN and CLAHE with YOLOv9 and EasyOCR has demonstrated a strong performance improvement in ANPR implementations. YOLOv9s outperformed other models by achieving the highest mAP@50 value of 99.3%, along with high precision and mAP50-95 scores. The smaller models offer a more balanced trade-off between accuracy and computational efficiency. This makes them well-suited for real-time ANPR applications in Indonesia. Furthermore, EasyOCR's accuracy in detecting Indonesian vehicle license plates has improved with the implementation of YOLOv9 and image enhancement techniques. Based on the 100 samples, EasyOCR was able to achieve the highest average accuracy of 84.36% with the proposed method. This means that the combination of Real-ESRGAN and CLAHE was able to address common challenges in real-world ANPR by reducing the effects of low resolution, blurring, and poor contrast.

However, extremely low-contrast visuals are still a challenge for YOLOv9 and EasyOCR. Image enhancement should be applied before the object detection step to guarantee that license plate objects can be detected even in low-quality images. One method to consider is automatic CLAHE with dual gamma correction. This method combines CLAHE with a two-level gamma correction technique and strict parameter control. The first gamma increases the overall image block brightness, whereas the second gamma adjusts the contrast in the darker areas. It is used to adaptively enhance contrast with strong dark shadows in each image block. Other techniques like denoising and sharpening could also make the characters on the license plate more clearly visible. Future studies will investigate additional image enhancing methodologies, including noise reduction approaches, to increase the quality of license plate images.

ACKNOWLEDGMENT

This paper is supported by the National Science and Technology Council, Taiwan (Grant No.: MOST-111-2637-H-324-001-). This research is partially supported by the Vice-Rector of Research, Innovation, and Entrepreneurship at Satya Wacana Christian University.

REFERENCES

- [1] Endarti, E., Wasiaturrahma, W., Sunyoto, K., Ibrahim, K.H., Putra, A.P. (2024). A critical review of land transport management in Indonesia: Does it support CO₂ emission reduction? *Social Sciences & Humanities Open Preprint*. <https://doi.org/10.2139/ssrn.4944095>
- [2] Indriani, M. (2022). Efektivitas penerapan e-tilang di Indonesia. *Khatulistiwa Profesional: Jurnal Pengembangan SDM dan Kebijakan Publik*, 3(1): 53-61. <https://doi.org/10.62099/khaprov.3i1.24>
- [3] Pangestu, I.D., Noor, H.T., Gunawan, G. (2023). Faktor utama penyebab pelanggaran lalu lintas oleh pengendara sepeda motor. *Jurnal Justitia: Jurnal Ilmu Hukum Dan Humaniora*, 6(2): 648-658. <https://doi.org/10.31604/justitia.v6i2.648-658>
- [4] Yuliantoro, Y., Sulchan, A. (2021). The effectiveness against traffic violations with electronic traffic law enforcement (ETLE). *Law Development Journal*, 3(4): 736-742. <https://doi.org/10.30659/ldj.3.4.736-742>
- [5] Nugroho, A.S. (2022). Electronic traffic law enforcement (ETLE) mobile sebagai difusi inovasi, interoperabilitas menuju etle nasioanl, (Studi Implementasi ETLE Mobile di Wilayah Provinsi Jawa Tengah). *Jurnal Ilmu Kepolisian*, 16(3): 157-176. <https://doi.org/10.35879/jik.v16i3.358>
- [6] Kumar, S., Sharma, M., Handa, K., Jaiswal, R. (2020). Improve OCR accuracy with advanced image preprocessing using machine learning with python. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 9(7): 1026-1030. <https://doi.org/10.35940/ijitee.g5745.059720>
- [7] Reddy, G.V., Gandla, A., Veldurthi, J., Dugyala, R., Mohammed, K.A., Swathi, B., Sobti, R., Kumar, A. (2024). A comprehensive review on helmet detection and number plate recognition approaches. In *E3S Web of Conferences*. EDP Sciences, 507: 01075. <https://doi.org/10.1051/e3sconf/202450701075>
- [8] Imran, A., Hulikal, M.S., Gardi, H.A. (2024). Real time American sign language detection using YOLO-v9. *arXiv Preprint arXiv: 2407.17950*. <https://doi.org/10.48550/arXiv.2407.17950>
- [9] Terven, J., Córdova-Esparza, D.M., Romero-González, J.A. (2023). A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, 5(4): 1680-1716. <https://doi.org/10.3390/make5040083>
- [10] Sarhan, A., Abdel-Rahem, R., Darwish, B., Abou-Attia, A., Sneed, A., Hatem, S., Badran, A., Ramadan, M. (2024). Egyptian car plate recognition based on YOLOv8, Easy-OCR, and CNN. *Journal of Electrical Systems and Information Technology*, 11(1): 32. <https://doi.org/10.1186/s43067-024-00156-y>
- [11] Tran-Anh, D., Tran, K.L., Vu, H.N. (2023). License plate recognition based on multi-angle view model. *arXiv Preprint arXiv: 2309.12972*. <https://doi.org/10.48550/arXiv.2309.12972>
- [12] Wang, C.Y., Yeh, I.H., Mark Liao, H.Y. (2024). YOLOv9: Learning what you want to learn using programmable gradient information. In *18th European Conference on Computer Vision, Milan, Italy*, pp. 1-21. https://doi.org/10.1007/978-3-031-72751-1_1
- [13] Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J. (2024). YOLOv10: Real-Time end-to-end object detection. *Advances in Neural Information Processing Systems*, 37: 107984-108011. <https://doi.org/10.48550/arXiv.2405.14458>
- [14] Wang, C.Y., Mark Liao, H.Y., Wu, Y.H., Chen, P.Y., Hsieh, J.W., Yeh, I.H. (2020). CSPNet: A New Backbone that can enhance learning capability of CNN. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Seattle, WA, USA, pp. 1571-1580. <https://doi.org/10.1109/CVPRW50498.2020.00203>
- [15] Wang, C.Y., Liao, H.Y.M., Yeh, I.H. (2022). Designing network design strategies through gradient path analysis. *arXiv Preprint arXiv: 2211.04800*.

- <https://doi.org/10.48550/arXiv.2211.04800>
- [16] Nazeem, M., Anitha, R., Navaneeth, S., Rajeev, R.R. (2024). Open-source OCR libraries: A comprehensive study for low resource language. In Proceedings of the 21st International Conference on Natural Language Processing (ICON), Chennai, India, pp. 416-421. <https://aclanthology.org/2024.icon-1.48/>.
- [17] Baek, Y., Lee, B., Han, D., Yun, S., Lee, H. (2019). Character region awareness for text detection. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, pp. 9365-9374. <https://doi.org/10.1109/CVPR.2019.00959>
- [18] Shi, B., Bai, X., Yao, C. (2016). An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11): 2298-2304. <https://doi.org/10.1109/TPAMI.2016.2646371>
- [19] Vedhaviyash, D.R., Sudhan, R., Saranya, G., Safa, M., Arun, D. (2022). Comparative analysis of EasyOCR and TesseractOCR for automatic license plate recognition using deep learning algorithm. In 2022 6th International Conference on Electronics, Communication and Aerospace Technology, Coimbatore, India, pp. 966-971. <https://doi.org/10.1109/ICECA55336.2022.10009215>
- [20] Kabiraj, A., Pal, D., Ganguly, D., Chatterjee, K., Roy, S. (2023). Number plate recognition from enhanced super-resolution using generative adversarial network. *Multimedia Tools and Applications*, 82(9): 13837-13853. <https://doi.org/10.1007/s11042-022-14018-0>
- [21] Wang, H., Li, Y., Dang, L.M., Moon, H. (2021). Robust Korean license plate recognition based on deep neural networks. *Sensors*, 21(12): 4140. <https://doi.org/10.3390/s21124140>
- [22] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Change Loy, C. (2018). ESRGAN: Enhanced super-Resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, pp. 63-79. https://doi.org/10.1007/978-3-030-11021-5_5
- [23] Wang, X., Xie, L., Dong, C., Shan, Y. (2021). Real-Esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF International Conference on Computer Vision. Montreal, BC, Canada, pp. 1905-1914. <https://doi.org/10.1109/ICCVW54120.2021.00217>
- [24] Schonfeld, E., Schiele, B., Khoreva, A. (2020). A U-Net based discriminator for generative adversarial networks. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, pp. 8207-8216. <https://doi.org/10.1109/CVPR42600.2020.00823>
- [25] Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y. (2018). Spectral normalization for generative adversarial networks. *arXiv Preprint arXiv: 1802.05957*. <https://doi.org/10.48550/arXiv.1802.05957>
- [26] Zuiderveld, K. (1994). Contrast limited adaptive histogram equalization. In *Graphics Gems IV*. Boston, MA, USA: Academic Press, pp. 474-485. <https://doi.org/10.1016/C2013-0-07360-4>
- [27] Dewi, C., Chen, R.C., Zhuang, Y.C., Manongga, W.E. (2024). Image enhancement method utilizing YOLO models to recognize road markings at night. *IEEE Access*, 12: 131065-131081. <https://doi.org/10.1109/ACCESS.2024.3440253>
- [28] Chang, Y., Jung, C., Ke, P., Song, H., Hwang, J. (2018). Automatic contrast-limited adaptive histogram equalization with dual gamma correction. *IEEE Access*, 6: 11782-11792. <https://doi.org/10.1109/ACCESS.2018.2797872>
- [29] Fan, R., Li, X., Lee, S., Li, T., Zhang, H.L. (2020). Smart image enhancement using CLAHE based on an F-shift transformation during decompression. *Electronics*, 9(9): 1374. <https://doi.org/10.3390/electronics9091374>
- [30] Pratomo, A.H., Kaswidjanti, W., Mu'arifah, S. (2020). Implementasi algoritma region of interest (ROI) untuk meningkatkan performa algoritma deteksi dan klasifikasi kendaraan. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 7(1): 155-162. <https://doi.org/10.25126/jtiik.202071718>
- [31] Singh, R., Goyal, S., Agarwal, S., Upadhyay, S. (2024). Evaluating the performance of ensemble YOLOv8 variants in smart parking applications for vehicle detection and license plate recognition under varying lighting conditions. *Preprints*. <https://doi.org/10.20944/preprints202404.1216.v1>
- [32] Amin, A., Mumtaz, R., Bashir, M.J., Zaidi, S.M.H. (2023). Next-generation license plate detection and recognition system using YOLOv8. In 2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET), Boca Raton, FL, USA, pp. 179-184. <https://doi.org/10.1109/HONET59747.2023.10374756>
- [33] Duda, R.O., Hart, P.E. (1972). Use of the Hough transformation to detect lines and curves in pictures. *Communications of the ACM*, 15(1): 11-15. <https://doi.org/10.1145/361237.361242>
- [34] Kim, S., Kim, M. (2023). Rotation representations and their conversions. *IEEE Access*, 11: 6682-6699. <https://doi.org/10.1109/ACCESS.2023.3237864>
- [35] Otsu, N. (1975). A threshold selection method from gray-level histograms. *Automatica*, 9(1): 62-66. <https://doi.org/10.1109/TSMC.1979.4310076>
- [36] Khairnar, S., Thepade, S.D., Gite, S. (2021). Effect of image binarization thresholds on breast cancer identification in mammography images using OTSU, Niblack, Burnsen, Thepade's SBTC. *Intelligent Systems with Applications*, 10: 200046. <https://doi.org/10.1016/j.iswa.2021.200046>
- [37] Levenshtein, V.I. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10(8): 707-710.
- [38] Final Project. Indonesian LPR final projectt computer vision project. *Roboflow*. <https://universe.roboflow.com/final-project-c3tpi/indonesian-lpr-final-projectt>, accessed on Oct. 7, 2024.