LPLC: A Dataset for License Plate Legibility Classification

Lucas Wojcik*, Gabriel E. Lima*, Valfride Nascimento*, Eduil Nascimento Jr.†, Rayson Laroca‡,*, David Menotti*
*Department of Informatics, Federal University of Paraná, Curitiba, Brazil
†Department of Technological Development and Quality, Paraná Military Police, Curitiba, Brazil
‡Graduate Program in Informatics, Pontifical Catholic University of Paraná, Curitiba, Brazil
*{lmlwojcik,gelima,vwnascimento,menotti}@inf.ufpr.br †eduiljunior@pm.pr.gov.br ‡rayson@ppgia.pucpr.br

***Abstract*—Automatic License Plate Recognition (ALPR) faces a major challenge when dealing with illegible license plates (LPs). While reconstruction methods such as super-resolution (SR) have emerged, the core issue of recognizing these low-quality LPs remains unresolved. To optimize model performance and computational efficiency, image pre-processing should be applied selectively to cases that require enhanced legibility. To support research in this area, we introduce a novel dataset comprising 10,210 images of vehicles with 12,687 annotated LPs for legibility classification (the LPLC dataset). The images span a wide range of vehicle types, lighting conditions, and camera/image quality levels. We adopt a fine-grained annotation strategy that includes vehicle- and LP-level occlusions, four legibility categories (perfect, good, poor, and illegible), and character labels for three categories (excluding illegible LPs). As a benchmark, we propose a classification task using three image recognition networks to determine whether an LP image is good enough, requires super-resolution, or is completely unrecoverable. The overall F1 score, below 80% for all three baseline models (ViT, ResNet, and YOLO), together with the analyses of SR and LP recognition methods, highlights the difficulty of the task and reinforces the need for further research. The proposed dataset is publicly available at https://github.com/lmlwojcik/lplc-dataset.**

I. Introduction

Automatic License Plate Recognition (ALPR) plays a vital role in road surveillance by using automated systems to detect vehicles and recognize them based on their license plates (LPs) [1], [2]. In the deep learning era, key research tasks in this domain include LP detection and LP recognition, often complemented by additional processes such as vehicle detection and LP rectification [3], [4].

State-of-the-art ALPR methods achieve over 95% accuracy in detection tasks and exceed 90% recognition rates under ideal conditions [5], [6]. However, their performance degrades significantly in challenging scenarios that are not well represented in most mainstream datasets. Such scenarios include low-light environments, adverse weather conditions (e.g., rain), and low-resolution or low-quality images caused by poor equipment or transmission compression artifacts [7], [8].

Furthermore, recent advances in super-resolution (SR) technology have also opened up new paths to deal with faulty images [9], [10]. These models were adapted to an LP reconstruction task and were shown to aid in recovering information from otherwise illegible images. However, these models are often costly and not always needed, potentially even ruining otherwise suitable images for LP recognition (as we have observed in some of our experiments reported in this work).

Fig. 1. Examples illustrating the distinction between image quality and LP legibility. High-quality images may contain illegible LPs (top right), while low-quality images can still include legible ones (bottom right). A single image may also feature both legible and illegible LPs (left).

In this context, it is important to distinguish between image quality and character legibility. LP recognition performance depends not merely on image resolution or overall visual quality, but primarily on how legible the characters are. Fig. 1 illustrates this point: a visually high-quality image may contain illegible LPs due to factors such as camera distance, while a low-quality image may still feature clearly legible LPs. In some cases, a single image can include both legible and illegible LPs. We also observed that Optical Character Recognition (OCR) models often produce incorrect predictions with high confidence [11], particularly in cases where most characters are legible but one or two remain unclear (an issue also depicted in Fig. 1).

With these challenges in mind, we present the *License Plate Legibility Classification* (LPLC) dataset[1], which contains over 10k radar images collected from various locations across the Brazilian state of Paraná and includes more than 12k annotated LPs. The images exhibit a wide range of capture conditions, including varying lighting scenarios, vehicle types (motorcycles, cars, buses, and trucks), and image quality. A key contribution of this dataset is the assignment of each LP to one of four legibility levels: perfect, good, poor, or illegible. These labels provide a qualitative and inherently subjective assessment of LP legibility. For the first three categories, we also provide OCR annotations. The dataset includes both Brazilian

[1]The LPLC dataset is available at https://github.com/lmlwojcik/lplc-dataset

Accepted for presentation at SIBGRAPI 2025. The final published version will be available on IEEE Xplore.

and Mercosur LPs[2] and can be used as a benchmark for various ALPR-related tasks.

In this work, we also use LPLC for a series of LP legibility assessment tasks using established image classification models: ResNet [12], ViT [13] and YOLO-cls [14]. Additionally, we explore the use of SR by evaluating three state-of-the-art models [9], [10], [15] on our novel dataset. Building on these components, we design a recognition pipeline that incorporates a new decision step to determine whether an input image is sufficiently legible, requires further processing, or should be considered unrecoverable.

The remainder of this work is structured as follows. Section II provides an overview of state-of-the-art approaches in LP recognition, commonly used benchmark datasets, and key challenges. Section III introduces our proposed dataset, describing the images, LPs, and annotated attributes. Section IV describes our experimental setup for image legibility classification using established image processing models. Section V presents and discusses the results. Finally, Section VI concludes the paper.

II. Related Work

State-of-the-art advancements in ALPR, as in other machine learning research domains, are quantitatively assessed using public benchmark datasets [3], [5]. An example is CLPD [16], a dataset containing 1,200 images captured with various devices in mainland China. While it primarily features daylight scenes and passenger cars, it offers considerable variability in backgrounds, road types, capture devices, and image quality.

For the Brazilian context, the RodoSol-ALPR dataset [5] comprises 20,000 images captured at toll plazas, providing a realistic representation of actual operating conditions. It includes a wide range of lighting scenarios, vehicle types, and weather conditions. The authors [5] demonstrated that state-of-the-art methods perform well on older datasets featuring simpler scenarios, such as Caltech-Cars [17], but struggle on RodoSol-ALPR, revealing limitations tied to outdated benchmarks. While RodoSol-ALPR captures real situations, it is important to note that these toll plaza environments differ significantly from many surveillance camera settings, where camera quality and capture conditions are often much poorer and more variable.

Recent advances in LP recognition have shifted focus from achieving high performance on simpler datasets to addressing more specific challenges in the field. For example, Liu et al. [6] proposed an attention-based decoder for LP recognition, following the growing trend of integrating attention mechanisms. Their approach combines a CNN encoder for feature extraction with a contrastive learning strategy designed to differentiate both the position and class of each character. The use of attention is motivated by the fixed character positions in LPs of a given format, enabling more effective layout-aware feature learning. They reported a 96% recognition rate on RodoSol-ALPR.

Similarly, Rao et al. [18] proposed an end-to-end pipeline that includes a segmentation step. The method begins with a YOLOv5-based LP detector, followed by segmentation using the proposed AFF-Net model. Next, a skew correction module is applied to the segmentation map using standard image processing techniques to extract the LP region. The four corners of this region are then aligned with those of a canonical LP format. Finally, LP recognition is performed using a CNN. This correction step is effective in handling severely skewed LPs captured from oblique angles, such as those found in the CLPD dataset, where they reported a recognition rate of 94%.

SR has gained interest in ALPR research due to its potential to recover otherwise illegible LP images. For example, studies like [10] adapt image-to-image translation methods, such as Generative Adversarial Network (GAN), to improve recognition performance on low-resolution LPs.

In the same direction, Nascimento et al. [9] proposed a novel focal loss tailored for character reconstruction, trained on a variant of the RodoSol-ALPR dataset called RodoSol-SR. This dataset comprises pairs of high-resolution LP images and their synthetically degraded low-resolution counterparts. These pairs were generated by the authors to train the model to reverse the degradation process through reconstruction.

Building on this context, we introduce the LPLC dataset, detailed in the following section. Each LP is annotated with a qualitative legibility score using four distinct levels, which, to the best of our knowledge, is a novel contribution to the literature. The dataset supports the development and evaluation of LP legibility classification methods, an important step in determining whether an image requires further processing such as super-resolution. It also serves as a more challenging benchmark for conventional LP detection and recognition tasks.

III. The LPLC Dataset

We introduce the *License Plate Legibility Classification* (LPLC) dataset, a publicly available resource primarily designed for the task of classifying the legibility of LP images, that is, whether an LP image is suitable for direct OCR or requires additional processing. Although its main purpose is legibility classification, the dataset is also suitable for other tasks such as LP detection and recognition, due to its fine-grained annotations that include details on LP text, legibility levels, and occlusion. All annotations were made manually by a single person (the first author) using the VIA annotation software [19], and validated through a semi-automatic process to ensure accuracy. This revision process involved using an OCR model's outputs (PARSeq-tiny [20]) and identifying legibility classification errors to flag potentially inconsistent annotations in both the legibility labels and OCR transcripts. Fig. 2 presents a few sample images from the dataset.

LPLC is composed of 10,210 images captured by traffic radars across the Brazilian state of Paraná. All images were processed and redacted to remove metadata embedded by the cameras. These images were taken from hundreds of different cameras on various roads around the covered area. Each image is annotated with the time of capture (morning, afternoon, evening, or night) and may contain one or more LPs, totaling 12,687 annotated LPs. An LP is annotated if a significant part (roughly two-thirds)

[2] Following prior literature [2], [9], [11], we use the term "Brazilian" to refer to the LP layout used in Brazil prior to the adoption of the Mercosur layout.

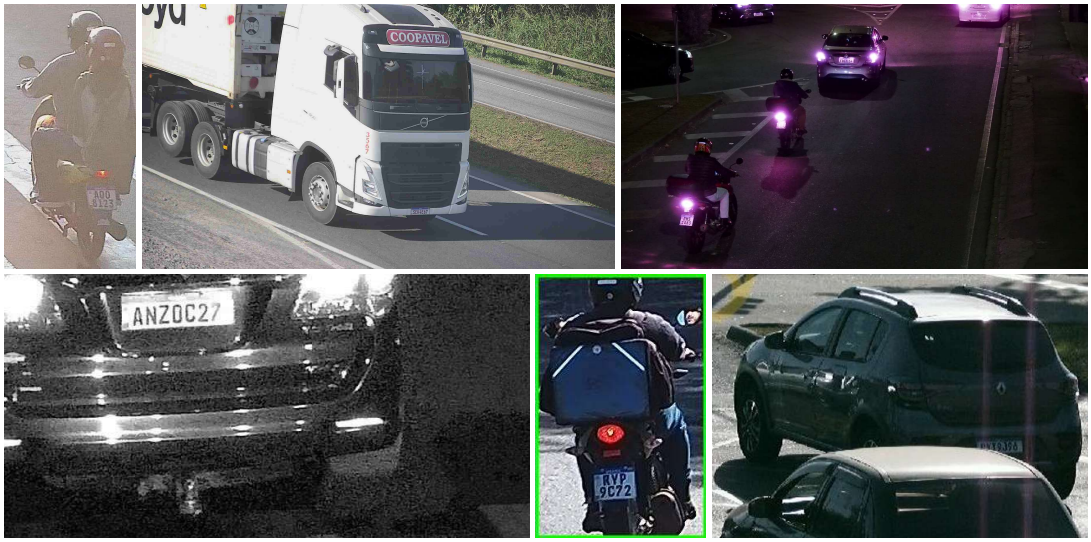

Fig. 2. Samples from the LPLC dataset.

of the car is present in the image and if the LP is large enough to assess legibility. Each LP is annotated with five attributes: (i) the coordinates of the LP, (ii) occlusion of the LP, (iii) occlusion of the vehicle, (iv) legibility level, and (v) the LP characters.

The coordinates correspond to the four corners of the LP as it appears in the image, which may form an irregular quadrilateral, starting at the top-left coordinate and moving clockwise towards the bottom-right. The LP-level occlusion is a binary attribute labeled "occluded," which is set to true if one or more characters on the LP are not visible. This often occurs when vehicles are positioned at the edge of the image but still meet our criteria for valid annotation, as seen in the bottom left vehicle in Fig. 2. Occlusion may also result from objects in the scene, such as other vehicles, tree branches, or lamp posts. This is illustrated in the bottom right vehicle in Fig. 2. Vehicle-level occlusion is indicated by a binary attribute labeled "valid." An LP is considered valid if it belongs to a vehicle with either most of its body (approximately 80% visible or the full front or rear view, including both headlights or taillights) captured in the image.

The annotation of LP characters is directly linked to their legibility level. All LPs were manually labeled and then cross-validated using the PARSeq-tiny OCR model [20]. LPs classified as illegible — the lowest legibility level — are assigned an empty string, as their text cannot be recovered. For the remaining classes, the characters were annotated and validated through a semi-automatic process: the OCR model was used to predict each LP's text, and any discrepancies between the manual and predicted annotations were reviewed and corrected.

The final validation step for potentially incorrect OCR annotations involved manually verifying whether the annotated LP text matched the vehicle shown in the image. This was done by querying a paid API to retrieve vehicle information — such as make, model, color, and year — based on the annotated LP characters. If the returned vehicle details did not match the one in the image, the annotation was reviewed and a new query was issued. This process was repeated until a correct match was found or up to four attempts had been made.

The legibility level is annotated based on the visual quality of the LP text in the image. We define four legibility levels, numbered from zero to three. Class 0, labeled "illegible," refers to LPs where the text is either completely unrecognizable or so degraded that it could not be validated using our method. Class 1, "poor," includes LPs with distorted text in which characters are not immediately recognizable. Class 2, "good," refers to LPs with legible text that may still exhibit some noise or visible distortion. Finally, Class 3, "perfect," corresponds to LPs with clearly visible characters and no noticeable distortion. Examples of each legibility level are shown in Fig. 3.

Fig. 3. OCR legibility levels.

Table I shows the distribution of LPs and attributes in the dataset. As indicated, the dataset is slightly imbalanced with respect to the legibility class. An OCR annotation is considered true if the LP's characters are labeled, even if some characters are missing due to occlusion. As previously noted, all LPs have annotated characters, except those categorized as illegible.

Table II presents the distribution of images in the dataset. While nighttime images constitute a minority of our dataset (4,665 evening and nighttime images versus 6,386 morning and afternoon images), this represents a significant contribution compared to other public datasets, which typically contain

TABLE I
LP Statistics.

LPs by Legibility		Other Attributes		
Class	Number	Class	True	False
Perfect	5,617	Occluded	12,586	101
Good	3,641	Valid	12,359	328
Poor	1,825	OCR	11,083	1,604
Illegible	1,604	→ Total LPs	12,687	

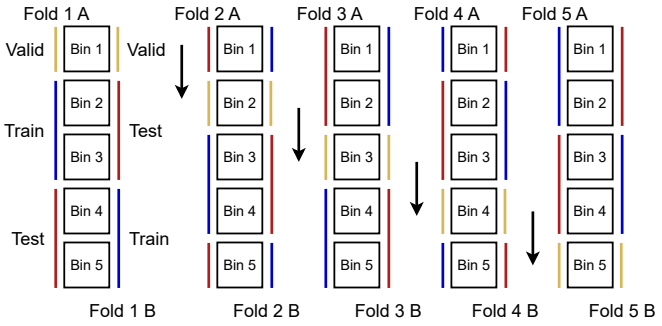

Fig. 4. Cross-fold Splits Illustration.

minimal nighttime imagery [21]. Most images contain at least one vehicle with a positive label (an LP that is legible, non-occluded, or associated with a valid vehicle), and 9,454 images include at least one LP that is legible, valid, and not occluded.

TABLE II
Image Statistics.

Images by Time of Day		Images by Attributes	
Class	Number	Has at Least One	Number
Morning	3,830	Legible LPs	9,684
Afternoon	2,556	Non Occluded LPs	10,195
Evening	2,585	Valid Vehicles	10,030
Night	1,239	→ Total Images	10,200

IV. Experimental Setup

Our experiments focus on the legibility attribute at the LP level. We crop all 12,687 LPs from the source images using their annotated corner coordinates. To evaluate legibility classification, we consider three models: ResNet [22], ViT [13] and YOLO-cls [14]. The goal is to assess whether an LP image is suitable for OCR processing, enabling more efficient downstream LP recognition by filtering out low-quality samples.

We employ a 5-fold cross-validation protocol with 40%, 20%, and 40% splits for training, validation, and testing, respectively. Each fold generates two different experiment iterations, where the training and test partitions are flipped, as illustrated in Fig. 4. Consequently, each experiment is carried out 10 times, once per unique fold configuration, and the final results are reported as the average over these 10 runs.

There are several reasons behind our choice of protocol. First, the n-fold cross-validation protocol allows us to use the entire dataset for testing, providing performance results for every image. Second, it mitigates bias associated with specific training and test partitions. Averaging results across all folds ensures that the evaluation is not influenced by a particularly easy or difficult split. Lastly, alternating the roles of training and test partitions between rounds allows for more accurate comparisons across folds. The fold splits are made available alongside the dataset.

We train each model under three scenarios, all framed as straightforward image classification tasks using the cropped LPs. The first scenario, called "Baseline," involves predicting one of the four standard legibility classes: perfect, good, poor, or illegible. This setup aims to evaluate how effectively current models can assess the legibility of a given LP image.

In the second scenario, termed "Legibility Recognition", we merge the *perfect* and *good* LPs into a single class "legible" and train the network to perform a binary classification between legible and poor LPs. *Illegible* LPs are excluded from this experiment. The objective is to determine whether an image is suitable for OCR or requires further processing — hence the binary output. The third scenario, "Full Recognition", builds on the second by reintroducing *illegible* LPs as a third class, framing the task as ternary classification. In addition to identifying images that may or may not require enhancement for OCR, this scenario also aims to detect unrecoverable images that should be discarded.

We also assess the impact of SR on LP recognition performance. For SR, we adopt the pipelines proposed in [9], where images are first reconstructed using their SR model and then processed with a version of the GP_LPR OCR model [23] trained on the RodoSol-ALPR dataset [5]. Both SR and OCR are applied to all images, and the OCR results from the reconstructed images are compared against those from the originals. In this experiment, in addition to the LCOFL-GAN proposed in [9], we also evaluate Real-ESRGAN [15] and LPSRGAN [10], providing a broader overview of current SR state of the art.

Finally, we report OCR results by legibility class. In addition to GP_LPR, we also evaluate the PARSeq-tiny model [20], trained on an unpublished dataset. We report both character-level and LP-level recognition rate, presenting the results broken down by legibility class.

A. Fine-Tuning Parameters

As mentioned before, we train ResNet, YOLO-cls, and ViT in these scenarios. Specifically, we employ ResNet-50 and ViT Base-16 from PyTorch's torchvision library, and YOLO11m-cls from the ultralytics implementation. These models were selected based on their proven effectiveness and widespread adoption in state-of-the-art research across diverse domains [24]–[26]. We initialize all models with the weights pre-trained on ImageNet, and swap the last layer for a simple linear layer with N outputs, where N is the number of classes in a given training scenario (4, 2, and 3 classes for scenarios 1, 2, and 3, respectively).

All experiments are conducted on an NVIDIA GeForce 3090 RTX GPU. Each model is trained across ten folds for a maximum of 200 epochs with a batch size of 16. We employ an early stopping strategy based on the validation set accuracy, with a patience of 20 epochs. For ViT and ResNet, we use the Adam optimizer at a learning rate of 10^{-5}. YOLO-cls is trained with

an SGD optimizer, a starting learning rate of 10^{-2}, multiplied by 10^{-2} following the cosine weight decay. For all models, we fine-tune every layer instead of only the final classifier layer.

V. Results

Table III presents our results on the *Baseline* scenario, which consists of predicting the legibility label for a given LP image. We report the average test micro-F1 score across the 10 folds for each class, along with the overall F1 score. Despite the task's straightforward nature and limited number of classes, the results reveal it to be quite challenging. Among the models evaluated, YOLO-cls achieves the best performance. Fig. 5 shows the confusion matrix of YOLO-cls for one of the folds, revealing substantial overlap between adjacent classes and blurred decision boundaries — a pattern that corresponds to the ambiguous cases where the model made the most errors.

TABLE III
Results on the Baseline scenario (F1-score).

Model	Class				Overall
	Perfect	Good	Poor	Illegible	
ResNet-50	84.54%	67.98%	56.70%	72.97%	74.51%
ViT b-16	85.74%	68.00%	58.80%	73.67%	75.48%
YOLO11m-cls	88.37%	65.83%	59.42%	74.47%	76.79%

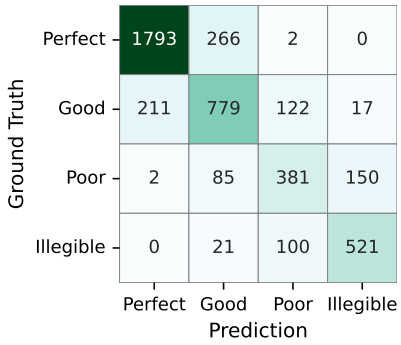

Fig. 5. Confusion matrix for one run of YOLO-cls.

Table IV presents the classification results for the *Legibility Recognition* and *Full Recognition* scenarios. The first scenario evaluates whether an LP image is suitable for OCR, while the second introduces a third class to identify LPs with unrecoverable text. We report the test micro-F1 score averaged across the 10 folds. Due to the reduced number of classes, this becomes a simpler task in which the model distinguishes between legible and non-legible (or unrecoverable) images. As expected, the classification metrics are correspondingly higher. Nevertheless, there remains considerable room for improvement, as the models consistently struggle to differentiate between high- and low-quality textures across all scenarios.

Table V presents the performance of the pre-trained SR models when integrated with LP recognition. All cropped LPs from the LPLC dataset were first enhanced using the respective SR model (LCOFL-GAN [9], RealESRGAN [15], and LPSRGAN [10]) and then processed with the GP_LPR OCR model [23], trained on the RodoSol-ALPR dataset [5] (as in [9]). This setup evaluates whether SR improves LP recognition accuracy. The results suggest otherwise: in the best case (Real-ESRGAN), recognition improved for only 647 out of the 11,027 LPs ($\leq 6\%$). These findings indicate that current SR networks struggle to generalize across datasets.

TABLE IV
Results on the Legibility scenario (F1-score).

Model	Legibility Recognition (Legible vs. Poor)	Full Recognition (Legible, Poor, Illegible)
ResNet-50	92.56%	87.23%
ViT b-16	93.16%	87.78%
YOLO11m-cls	92.71%	86.25%

TABLE V
SR network evaluation (character accuracy).

GAN Model	OCR Results With SR	Class			Total
		Perfect	Good	Poor	
LCOFL-GAN [9]	Better	98	108	108	314
	Equal	494	308	170	972
	Worse	5,012	3,206	1,523	9,741
Real-ESRGAN [15]	Better	211	276	160	647
	Equal	5,180	2,683	572	8,435
	Worse	213	663	1,069	1,945
LPSRGAN [10]	Better	108	98	45	251
	Equal	264	209	115	588
	Worse	5,232	3,315	1,641	10,188

This demonstrates that applying super-resolution does not necessarily enhance LP recognition performance and can, in many cases, be detrimental. Fig. 6 shows six reconstructed images (using LCOFL-GAN) alongside their original versions, organized by legibility level. As can be seen, the employed SR model not only reduces legibility but also introduces hallucinated characters that were not present in the original images.

Fig. 6. Original and reconstructed images using super-resolution.

Finally, Table VI presents the character-wise accuracy (Char Acc) and whole-LP accuracy (LP Acc) for our dataset, broken down by legibility class. As described in Section IV, we use GP_LPR [23] and PARSeq-tiny [20], the latter of which was trained on a different, unpublished dataset. The results clearly indicate that OCR performance is strongly correlated with the legibility of the LP: lower legibility levels lead to significantly reduced recognition accuracy.

TABLE VI
OCR EVALUATION ON THE LPLC DATASET.

Class	GP_LPR [23]		PARSeq-tiny [20]	
	Char Acc	LP Acc	Char Acc	LP Acc
Perfect	90.57%	80.74%	99.52%	98.07%
Good	85.02%	62.37%	98.40%	92.32%
Poor	74.91%	30.30%	93.78%	72.34%
Overall	86.30%	66.84%	98.08%	91.37%

VI. Conclusions

In this work, we introduced a novel ALPR-related dataset consisting of 10,200 images captured by street radars, encompassing a diverse range of vehicles and LPs. The dataset includes fine-grained annotations such as legibility level, LP text, and both vehicle- and LP-level occlusion. We adapted three image classification models for the task of legibility classification and evaluated them in a 5-fold cross-validation setup, reporting average results across ten runs.

Our findings indicate that although the evaluated classification networks achieve promising results, they are not yet reliable for deployment in real-world applications, particularly due to their difficulty in distinguishing subtle legibility differences. Furthermore, the super-resolution model evaluated in this study exhibits poor generalization in cross-dataset settings, frequently degrading LP legibility or introducing hallucinated characters, ultimately harming LP recognition performance.

As future work, we suggest the development of SR models tailored for cross-domain scenarios, as well as the exploration of alternative image enhancement techniques aimed at improving LP legibility without compromising the integrity of the visual content. In addition, we plan to position this dataset within the state of the art by conducting a more comprehensive comparative evaluation against current ALPR methods, thereby demonstrating its broad applicability.

Acknowledgments

This study was financed in part by the [ORGANIZATION 1], and in part by the [ORGANIZATION 2]. We gratefully acknowledge the support of [CORPORATION] with the donation of [EQUIPMENT] used for this research.

References

[1] R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, and D. Menotti, "An efficient and layout-independent automatic license plate recognition system based on the YOLO detector," *IET Intelligent Transport Systems*, vol. 15, no. 4, pp. 483–503, 2021.

[2] S. M. Silva and C. R. Jung, "A flexible approach for automatic license plate recognition in unconstrained scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5693–5703, 2022.

[3] X. Ke, G. Zeng, and W. Guo, "An ultra-fast automatic license plate recognition approach for unconstrained scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 5, pp. 5172–5185, 2023.

[4] H. Ding, J. Gao, Y. Yuan, and Q. Wang, "An end-to-end contrastive license plate detector," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 1, pp. 503–516, 2024.

[5] R. Laroca *et al.*, "On the cross-dataset generalization in license plate recognition," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, Feb 2022, pp. 166–178.

[6] Q. Liu, Y. Liu, S.-L. Chen, T.-H. Zhang, F. Chen, and X.-C. Yin, "Improving multi-type license plate recognition via learning globally and contrastively," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 9, pp. 11 092–11 102, 2024.

[7] V. Wahyu Saputra, N. Suciati, and C. Fatichah, "Fog and rain augmentation for license plate recognition in tropical country environment," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 13, no. 4, p. 3951, Dec. 2024.

[8] V. Nascimento *et al.*, "Toward advancing license plate super-resolution in real-world scenarios: A dataset and benchmark," *Journal of the Brazilian Computer Society*, vol. 1, no. 31, pp. 435–449, 2025.

[9] ——, "Enhancing license plate super-resolution: A layout-aware and character-driven approach," *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 1–6, 2024.

[10] Y. Pan, J. Tang, and T. Tjahjadi, "LPSRGAN: Generative adversarial networks for super-resolution of license plate image," *Neurocomputing*, vol. 580, p. 127426, 2024.

[11] R. Laroca, L. A. Zanlorensi, V. Estevam, R. Minetto, and D. Menotti, "Leveraging model fusion for improved license plate recognition," in *Iberoamerican Congress on Pattern Recognition*, Nov 2023, pp. 60–75.

[12] T. S. Prajwal and I. A K, "A comparative study of resnet-pretrained models for computer vision," in *2023 Fifteenth ACM International Conference on Contemporary Computing*, 2023, p. 419–425.

[13] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021, pp. 1–22.

[14] Ultralytics, "YOLOv11," 2025, accessed: 2025-08-03. [Online]. Available: https://docs.ultralytics.com/models/yolo11/

[15] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," in *International Conference on Computer Vision Workshops (ICCVW)*, 2021.

[16] L. Zhang, P. Wang, H. Li, Z. Li, C. Shen, and Y. Zhang, "A robust attentional framework for license plate recognition in the wild," *Trans. Intell. Transport. Sys.*, vol. 22, no. 11, p. 6967–6976, Nov. 2021.

[17] M. Weber and P. Perona, "Caltech cars 1999," 2022. [Online]. Available: https://data.caltech.edu/records/20084

[18] Z. Rao, D. Yang, N. Chen, and J. Liu, "License plate recognition system in unconstrained scenes via a new image correction scheme and improved CRNN," *Expert Systems with Applications*, vol. 243, p. 122878, 2024.

[19] A. Dutta and A. Zisserman, "The VIA annotation software for images, audio and video," in *Proceedings of the 27th ACM International Conference on Multimedia*, ser. MM '19. New York, NY, USA: ACM, 2019. [Online]. Available: https://doi.org/10.1145/3343031.3350535

[20] D. Bautista and R. Atienza, "Scene text recognition with permuted autoregressive sequence models," in *European Conference on Computer Vision (ECCV)*, 10 2022, pp. 178–196.

[21] M. S. H. Onim, H. Nyeem, M. W. Khan Arnob, and A. D. Pooja, "Unleashing the power of generative adversarial networks: A novel machine learning approach for vehicle detection and localisation in the dark," *Cognitive Computation and Systems*, vol. 5, pp. 169–180, 2023.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[23] Y.-Y. Liu, Q. Liu, S.-L. Chen, F. Chen, and X.-C. Yin, "Irregular license plate recognition via global information integration," in *International Conference on Multimedia Modeling*, 2024, pp. 325–339.

[24] A. Vijayakumar and S. Vairavasundaram, "YOLO-based object detection models: A review and its applications," *Multimedia Tools and Applications*, vol. 83, no. 35, p. 83535–83574, Mar. 2024.

[25] G. E. Lima *et al.*, "Toward enhancing vehicle color recognition in adverse conditions: A dataset and benchmark," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Sept 2024, pp. 1–6.

[26] R. Laroca, M. dos Santos, and D. Menotti, "Improving small drone detection through multi-scale processing and data augmentation," in *International Joint Conference on Neural Networks (IJCNN)*, 2025.