Enhancing License Plate Super-Resolution: A Layout-Aware and Character-Driven Approach

Valfride Nascimento*, Rayson Laroca†,*, Rafael O. Ribeiro‡, William Robson Schwartz§, David Menotti*

*Federal University of Paraná, Curitiba, Brazil †Pontifical Catholic University of Paraná, Curitiba, Brazil
‡Brazilian Federal Police, Brasília, Brazil §Federal University of Minas Gerais, Belo Horizonte, Brazil

*{vwnascimento,menotti}@inf.ufpr.br †rayson@ppgia.pucpr.br ‡rafael.ror@pf.gov.br §william@dcc.ufmg.br

***Abstract*—Despite significant advancements in License Plate Recognition (LPR) through deep learning, most improvements rely on high-resolution images with clear characters. This scenario does not reflect real-world conditions where traffic surveillance often captures low-resolution and blurry images. Under these conditions, characters tend to blend with the background or neighboring characters, making accurate LPR challenging. To address this issue, we introduce a novel loss function, Layout and Character Oriented Focal Loss (LCOFL), which considers factors such as resolution, texture, and structural details, as well as the performance of the LPR task itself. We enhance character feature learning using deformable convolutions and shared weights in an attention module and employ a GAN-based training approach with an Optical Character Recognition (OCR) model as the discriminator to guide the super-resolution process. Our experimental results show significant improvements in character reconstruction quality, outperforming two state-of-the-art methods in both quantitative and qualitative measures. Our code is publicly available at https://github.com/valfride/lpsr-lacd.**

I. Introduction

Automatic License Plate Recognition (ALPR) systems have become increasingly popular in recent years, driven by their diverse practical applications, including toll collection, traffic monitoring, and forensic investigations [1]–[3].

These systems generally encompass two main tasks: License Plate Detection (LPD) and License Plate Recognition (LPR). LPD is concerned with locating the areas in an image that contain License Plates (LPs), whereas LPR is dedicated to identifying the characters on these LPs. Recent studies have been particularly concentrated on the LPR stage. Although the reported recognition rates have typically been high, most research has been conducted on high-resolution LPs, where the characters are easily discernible and clearly defined [4]–[6].

Surveillance cameras typically record images with low resolution or poor quality [7], mainly because of bandwidth and storage constraints. This leads to a scenario where LP characters blend with the background and adjacent characters, seriously affecting the effectiveness of LPR systems [8]–[10].

Various image enhancement techniques, including super-resolution and denoising methods, have been developed to improve image quality by increasing resolution or overall clarity. Despite significant strides in this field, most approaches focus on enhancing objective image quality metrics such as Peak Signal-to-Noise Ratio (PSNR) or Structural Similarity Index Measure (SSIM) without considering the specific application at hand [11], [12]. These techniques often struggle to differentiate between similar characters in low-resolution (LR) images, such as ‘B’ and ‘8’, ‘G’ and ‘6’, and ‘T’ and ‘7’.

In this work, we propose a perceptual loss function named *Layout and Character Oriented Focal Loss* (LCOFL) to enhance LP super-resolution and, consequently, LPR performance. LCOFL guides the network’s learning by considering not only factors such as resolution, texture, and structural details, but also the performance of the LPR task itself. It specifically penalizes errors related to character confusion and layout inconsistencies (i.e., LP sequences that deviate from the patterns observed in the training set), thus mitigating incorrect reconstructions. To further improve performance, we incorporate a Generative Adversarial Network (GAN)-style training, leveraging predictions from an Optical Character Recognition (OCR) model as the discriminator. Notably, LCOFL can utilize predictions from any OCR model since it relies solely on raw character predictions.

In summary, the main contributions of this work are:

- A novel loss function crafted to elevate the reconstruction of LP characters by integrating character recognition within the super-resolution process;
- Improvements to prevalent architectures in prior works by incorporating deformable convolution layers and shared weights into the attention module. A GAN-based training approach is also proposed, employing an OCR model as the discriminator. These strategies are aimed at generating LP images that are not only of high quality and resolution but also more accurately recognizable by OCR models;
- We have made our source code publicly available, aiming to stimulate further research within this domain.

II. Related Work

In this section, we review related works. First, we discuss approaches used for LPR in Section II-A. Then, we cover super-resolution methods for LPR in Section II-B.

A. License Plate Recognition

Silva & Jung [13] proposed to frame LPR as an object detection task with individual characters treated as unique classes for detection and recognition. They developed CR-NET, a YOLO-based model tailored specifically for LPR, which has subsequently proven highly effective and has been adopted in several follow-up studies [1], [4], [14].

Recent advancements have brought forth segmentation-free approaches using LPs to classify text from an entire LP image as a sequence. Ke et al. [15] developed a lightweight multi-scale LPR network with depthwise separable convolutional residual blocks and a multi-scale feature fusion layer for quick inference. Liu et al. [16] introduced deformable spatial attention modules to integrate global layout, enhancing character feature extraction. Rao et al. [6] tackled large-angle LP deflections caused by cameras by integrating channel attention mechanisms into CRNN, thereby improving LPR performance.

Multi-task models have also demonstrated considerable success for LPR. In these models, convolutional layers first process the entire LP image. Then, the network splits into N separate branches with fully connected layers. Each branch is dedicated to recognizing a single character (or blank), enabling the simultaneous prediction of up to N characters [17]–[19].

While these methods have shown considerable success in LPR, most of them were trained and evaluated using high-resolution (HR) images. However, this scenario does not mirror actual surveillance conditions, where LR images are prevalent. LR images often have characters that blend into the LP background due to quality issues and compression techniques used for storage [2], [9], [20].

B. Super-Resolution for License Plate Recognition

Recent advancements in deep learning have significantly improved general text super-resolution. However, there remains a notable gap in research dedicated to specifically enhancing LP recognition in low-resolution images.

Lin et al. [21] proposed using SRGAN [22] for LP super-resolution. Their approach was subsequently refined by Hamdi et al. [23], who introduced Double Generative Adversarial Networks (D_GAN_ESR_) for denoising, deblurring, and super-resolving LPs. However, these approaches lacked integration of character recognition within the learning process.

Focusing more on LP recognition, Pan et al. [24] proposed a pipeline for LP super-resolution and recognition, exploring ESRGAN [25] for single-character super-resolution. While effective, this method struggles with heavily degraded LPs where character boundaries are unclear. To enhance that approach, the same authors [10] later introduced a new degradation model to simulate low-resolution LPs and proposed LPSRGAN, an extension of ESRGAN that processes entire LP images. LPSRGAN, trained with an OCR-based loss function, better preserves character details and LP structural features but generates incorrect text in complex degradation cases [10].

Kim et al. [26] developed AFA-Net, combining super-resolution with pixel and feature-level deblurring to address motion blur in LPs. Although their findings showed potential, their test methodology was somewhat oversimplified, featuring LR images that remain legible and focusing solely on super-resolving digits while excluding letters and Korean characters.

Lee et al. [27] devised a GAN-based model incorporating a character perceptual loss utilizing features from ASTER, a well-known scene text recognition model. More recently, Nascimento et al. [9] introduced subpixel-convolution layers and an attention module to enhance textural and structural details, also with a perceptual loss integrating an OCR model. Despite their achievements, both methods encountered difficulties related to character confusion caused by structural or font similarities, resulting in reconstructions that deviated from the expected patterns in the specific LP layout under investigation.

III. Proposed Approach

This section details the proposed methodology. In Section III-A, we introduce our novel perceptual loss function, focusing on enhancing character reconstruction while accounting for the LP layout. In Section III-B, we outline the model's architecture, expanding on the framework introduced in [9], where it achieved significant success.

A. Layout and Character Oriented Focal Loss

In low-resolution LPs, characters often lose their shape details, blending with the LP's background and neighboring characters. Moreover, the lack of precise character positioning relative to the LP layout often causes the model to incorrectly super-resolve a letter as a digit or vice versa [8], [9], [28].

To enhance network guidance for LP reconstruction, we designed the Layout and Character Oriented Focal Loss (LCOFL) based on four key insights: (i) treating reconstruction partially as a classification task, where the super-resolved characters within an LP image need to be correctly identified by an OCR model; (ii) recognizing that characters typically adhere to specific patterns based on the LP layout (this includes fixed positions for digits and letters), which should be maintained during reconstruction; (iii) ensuring accurate reconstruction of characters with similar structures (e.g., "2" and "Z" or "R" and "B"); and (iv) preserving structural details from the original Ground Truth (GT) images in the final super-resolved images.

1) Classification Loss: To tackle the LP recognition problem, we adopt a weighted cross-entropy loss within the LCOFL function, as defined in Eq. (1):

$$L_C = -\frac{1}{K}\sum_{k=1}^{K} w_k \log p_t(y_k^{GT}|x_k)\,, \tag{1}$$

where $p_t(y_k^{GT})$ represents the predicted probabilities of the encoded GT label y_t^{GT} for the k-th character, while x represents the predicted probabilities for the super-resolved images. K is the maximum decoding length for the OCR alphabet.

The weights w are initialized as $[1, \ldots, K]$, where no penalization is applied, and each position corresponds to an encoded character in the OCR alphabet. These weights act as penalties assigned to characters misclassified due to structural or font similarities. Following the validation phase in each epoch, a confusion matrix is generated to assess the recognition of super-resolved images by the OCR model against the GT labels. Based on the confusion matrix, pairs of characters frequently confused with each other are identified, and an α value is added to the respective position of the encoded character in w, which is utilized during the training phase.

2) LP Layout Penalty: In various regions, including the one explored in this study (Brazil), digits and letters adhere to a fixed positional arrangement within the LPs. This means that a digit should not be mistakenly reconstructed as a letter, and vice versa. To enforce this constraint, a layout penalty, as defined in Eq. (2), is incorporated into the overall loss function.

$$L_P = \sum_{i=1}^{K} \left[D(x_k) \cdot A(y_k^{GT}) + \cdot A(x_k) \cdot D(y_k^{GT}) \right] \quad (2)$$

In Eq. (2), $D(\cdot)$ indicates the assertion of a digit at a specific position, while $A(\cdot)$ denotes the assertion of a letter. The left side of the sum verifies the correct positioning of a letter, while the right side checks the placement of a digit. As per the criteria outlined in Eq. (3), there is no penalty added if a character is correctly positioned. However, for each misplacement, a penalty value β is added to the sum.

$$\begin{aligned} \mathrm{D}(c) &= \begin{cases} \beta & \text{if } c \text{ is a digit} \\ 0 & \text{otherwise} \end{cases} \\ \mathrm{A}(c) &= \begin{cases} \beta & \text{if } c \text{ is a letter} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (3)$$

3) Dissimilarity Loss: Most objective methods for assessing image quality compare a reference, distortion-free image with a sample image. One commonly used metric is Mean Squared Error (MSE), which measures the average squared difference between each pixel of the reference and the sample image. MSE is also used in PSNR to calculate the ratio between the maximum possible value of a signal (the reference image) and its corrupted counterpart. These metrics are popular due to their simplicity and clear physical meaning [29].

However, they do not align well with the human visual system, which excels at identifying structural aspects of a scene, such as contrast, textures, structures, and luminance differences [30]. To better guide the network in super-resolving images with a focus on structural information, we incorporate SSIM [31] into our loss function, as shown in Eq. (4):

$$L_S = \frac{1 - SSIM(S_i, H_i)}{2}, \quad (4)$$

where S_i represents the super-resolved image, and H_i stands for the high-resolution GT image. SSIM evaluates three key aspects: luminance, contrast, and structure. SSIM values range from -1 to 1, with values near -1 indicating very different images and values near 1 indicating highly similar images in terms of structural similarity.

In Eq. (4), the transformation $(1 - SSIM)/2$ adjusts the SSIM values to a range of $[0, 1]$. This adjustment ensures that 0 represents highly similar images, while 1 represents highly dissimilar images. This adjustment facilitates its integration into the loss function, enabling more effective penalization of the network for generating images that deviate significantly from the GT in terms of structure.

After considering the classification loss (L_C), the LP layout penalty (L_P), and the dissimilarity loss (L_S), the final loss function is formulated in Eq. (5). The main goal is to improve the recognition rates attained by an OCR model when dealing with LR images. Simultaneously, it aims to restore intricate details to facilitate subsequent forensic analysis, which often represents the ultimate goal of the super-resolution process.

$$loss = L_D + L_P + L_S \quad (5)$$

B. Architecture

We extend the network architecture introduced in [9], specifically addressing the arrangement of Pixel Level Three-Fold Attention Modules (PLTFAMs) within the Residual Concatenation Blocks (RCBs) [11].

Nascimento et al. [9] focused on the design of the PLTFAM. Essentially, this design strategy capitalizes on *PixelShuffle*'s capabilities to utilize inter-channel feature relationships within the Channel Unit. It also integrates positional localization through the Positional Unit and improves texture and reconstruction with the Geometrical Perception Unit. Despite achieving remarkable recognition results, the attention module is uniquely defined in each RCB, which hampers the attention module's learning potential across individual RCBs.

To overcome this limitation, we propose incorporating an attention module with shared weights throughout the network structure. This enables the attention module to gather information from the initial layers, which extract fundamental LP characteristics, to the final layers, where finer details are restored. In this way, the attention module can consistently emphasize learning features essential for accurate LP recognition.

Furthermore, we have replaced the depthwise convolutions with deformable convolution layers in both the Positional and Channel units. Traditional convolution layers apply fixed geometric transformations, which may not adequately capture intricate, non-uniform deformations present in LR or degraded LP images. This limitation can restrict the network's capability to handle the diverse spatial arrangements of characters effectively. In contrast, deformable convolution layers dynamically adjust their receptive fields based on input features, allowing for more adaptable and precise modeling of spatial dependencies [3]. This adaptability significantly enhances the network's capacity to accurately reconstruct characters for recognition, even under challenging conditions. Consequently, the model's overall LPR performance is improved [3], [32].

IV. Experiments

This section covers our experiments, starting with the experimental setup. We then analyze the results, emphasizing significant improvements in recognition accuracy and reconstruction quality. Afterward, we conduct an ablation study to evaluate each component's contribution to the overall results. Finally, we present initial experiments conducted on real-world data.

A. Setup

We conducted our experiments using the RodoSol-ALPR dataset [33], which includes $10{,}000$ images of cars obtained at toll stations. This dataset includes $5{,}000$ images featuring cars with Brazilian LPs and another $5{,}000$ depicting cars

with Mercosur LPs[1]. Brazilian LPs are composed of 3 letters followed by 4 digits, while the initial pattern adopted for Mercosur LPs in Brazil comprises 3 letters, 1 digit, 1 letter, and 2 digits. We chose the RodoSol-ALPR dataset due to its diversity and frequent use in recent research [3], [5], [8]. In Fig. 1, we present LPs extracted from the dataset (cropped and rectified), showcasing differences in lighting, color combinations, and character fonts.

Fig. 1. Some LP images from the RodoSol-ALPR dataset [33]. The first two rows show Brazilian LPs, while the last two show Mercosur LPs. This work focuses on LPs with all characters arranged in a single row (10k images).

We explored the LR-HR pairs created and made available by Nascimento et al. [9]. The LR versions of each HR image were generated by simulating the effects of a low-resolution optical system with heavy degradation due to environmental factors or compression techniques. This was done by iteratively applying random Gaussian noise and resizing with bicubic interpolation until reaching the desired degradation level of SSIM < 0.1.

We also applied padding to the LR and HR images using gray pixels to preserve their aspect ratio before resizing them to 16×48 and 32×96 pixels, respectively, which corresponds to an upscale factor of 2. Fig. 2 shows examples of LP images obtained through this process.

Fig. 2. Examples of HR-LR image pairs used in our experiments.

In our GAN-based training methodology, we employed the OCR model proposed by Liu et al. [3] (GP_LPR) as the discriminator and the super-resolution model proposed in [9] (Pixel-Level Network (PLNET)) as the generator. During the testing phase, we employed the multi-task OCR model proposed by Gonçalves et al. [17], which has demonstrated significant success in previous studies [8], [9]. The decision to use different OCR models during training and testing was made to prevent biased reconstructions in the testing process.

The Adam optimizer was used with a learning rate of 10^{-4} for all models. To address oscillations during training, we implemented the StepLR learning rate schedule as recommended by Liu et al. [3]. This approach entails reducing the learning rate by a factor of 0.9 every 5 epochs if no improvement in recognition rate is observed. The GP_LPR model was set up with a decoding length of $K = 7$, corresponding to the 7-character format found in LPs in the RodoSol-ALPR dataset.

The experiments were performed using PyTorch framework on a computer with an AMD Ryzen 9 5950X CPU, 128 GB of RAM, and an NVIDIA Quadro RTX 8000 GPU (48 GB).

B. Experimental Results

LPR models are typically evaluated using recognition rates, defined as the number of correctly recognized LPs divided by the total number of LPs in the test set [4], [15], [19]. Given the prevalence of low-quality/low-resolution images in surveillance systems, we also report partial matches where at least six or at least five characters are correctly recognized. These partial matches are valuable in forensic applications because they can significantly narrow down the list of candidate LPs.

The results are presented in Table I. The first two rows show the recognition rates achieved by the OCR model [17] on the original HR images and their corresponding LR counterparts. This performance serves as a baseline to demonstrate the challenges faced by OCR models in low-resolution scenarios. As can be seen, the OCR model performs poorly on low-resolution LPs, achieving only a 1.1% recognition rate.

TABLE I
RECOGNITION RATES ACHIEVED ON DIFFERENT TEST IMAGES.

Test Images	# Correct Characters		
	All	≥ 6	≥ 5
HR (original images)	98.5%	99.9%	99.9%
LR (degraded images)	1.1%	5.3%	14.3%
LR + SR (PLNET [9])	39.0%	59.9%	74.2%
LR + SR (SR3 [35])	43.1%	67.5%	82.2%
LR + SR (**Proposed**)	**49.8%**	**71.2%**	**83.3%**

In the lower section of Table I, we present the OCR model's performance on LR images enhanced with super-resolution techniques. Our super-resolution network considerably outperformed two state-of-the-art baselines [9], [35]. Specifically, the OCR model showed notable performance improvements when the LPs were super-resolved by our model. It attains a recognition rate of 49.8%, in contrast to 43.1% using SR3 [35], a renowned diffusion method by Google Research, and 39.0% by PLNET [9]. Notably, our model took approximately 38 minutes to super-resolve all 4k test images (batch size = 1), whereas SR3 required around 33 hours ($52\times$ slower), thus underscoring our method's superior accuracy coupled with significantly reduced computational processing time. Naturally, to ensure a fair comparison, all models (i.e., our proposed approach and the baselines) were trained on the same samples.

In Fig. 3, we present a comparison of images enhanced by our method with those using the baselines [9], [35]. Remarkably, the proposed approach outperforms the baseline methods by accurately distinguishing between letters and digits, maintaining the original texture and structural details, and achieving superior visual quality in the reconstruction.

PLNET [9] performs quite well in reconstructing textures and structures but faces challenges with character confusion. For instance, it mistakenly reconstructed an "S" as a "5" in the

[1]Following prior literature [4], [8], [34], we use the term "Brazilian" to refer to the LP layout used in Brazil prior to the adoption of the Mercosur layout.

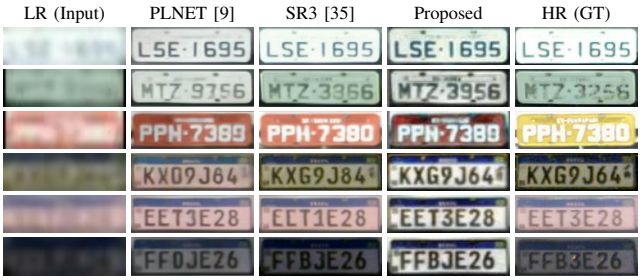

Fig. 3. Representative images produced by the proposed approach and baseline methods for the same inputs. GT = Ground Truth.

top row and a "G" as a "0" in the fourth row. In contrast, the SR3 model [35] accurately reconstructs characters according to the LP layout, reducing confusion between letters and digits. Nevertheless, SR3 still presents inconsistencies, such as partially reconstructing the letter "E" without its central line in the fifth row and exhibiting variations in the curvature of the two "F" in the bottom row.

The proposed approach showcases superior character reconstruction by consistently generating well-defined, super-resolved characters that adhere closely to the LP layout. This results in minimal discrepancies, particularly those arising from poor lighting conditions. For instance, in the final row of Fig. 3, while both baseline methods reconstructed a "J" due to a light spot, our approach aimed for a "3," aligning more accurately with the digit's structure. Furthermore, our method excels in maintaining consistent shapes and contours of characters across various LPs, leading to more reliable reconstructions. This is evident when analyzing the curves of the letter "E" and the two "F"s in the bottom two rows.

C. Ablation Study

This work introduces a novel loss function and incorporates several modifications into the training procedure, including a GAN-based style and changes to architectures proposed in previous works, particularly in [9]. To assess the contribution of each component, this section presents an ablation study.

We trained several networks under various conditions. These conditions included implementations with and without the proposed architectural modifications (ArchMod) detailed in Section III-B, with and without the GAN-based training style (GAN-style) described in Section IV-A, and utilizing either our proposed loss function (LCOFL) or the loss function from Nascimento et al. [9], which incorporates the logits from the OCR model proposed by Goncalves et al. [17]. In the experiments without the GAN-based style, predictions from the OCR model [17] on the super-resolved images served as input for the LCOFL loss (similar to [9]). These experiments led to seven baselines, and the results are detailed in Table II.

The ablation results demonstrate that each component contributes to the overall performance. Excluding the LCOFL leads to a significant performance decrease (from 49.8% to 45.9%). Removing ArchMod and GAN-style training also results in performance drops, albeit less severe, to 49.2% and 49.4%, respectively. Notably, excluding all the proposed components reduces the recognition rate drastically to just 39.0%.

TABLE II
RECOGNITION RATES (RR) ACHIEVED WITH DIFFERENT COMPONENTS INTEGRATED INTO THE PROPOSED APPROACH.

Approach	RR
Proposed (w/o ArchMod, GAN-style, and LCOFL)	39.0%
Proposed (w/o LCOFL)	45.9%
Proposed (w/o ArchMod and LCOFL)	47.6%
Proposed (w/o GAN-style and LCOFL)	47.7%
Proposed (w/o ArchMod and GAN-style)	48.2%
Proposed (w/o ArchMod)	49.2%
Proposed (w/o GAN-style)	49.4%
Proposed	**49.8%**

Based on these results, we argue that LCOFL plays a crucial role in aiding the network to accurately position the characters according to the LP layout and mitigate potential confusion caused by structural or font similarities among characters. Additionally, incorporating shared weights in PLTFAM and integrating deformable convolutions into its architecture have significantly enhanced the attention module's ability to extract structural and textural features relevant to the characters.

D. Preliminary Experiments on Real-World Data

We also explored a dataset of $3,723$ LR-HR image pairs collected from real-world settings. We allocated 80% of the pairs for training, 10% for validation, and 10% for testing. Table III shows the recognition rates achieved on the test images. The results reinforce the superiority of our super-resolution method, with the OCR model achieving a recognition rate of 39.5%, compared to 36.3% for PLNET [9] and 31.7% for SR3 [35]. Fig. 4 visually demonstrates the effectiveness of the proposed method. As one example, in the left image, both baseline methods incorrectly reconstructed an "O" instead of a "U," with PLNET also introducing noticeable artifacts. In contrast, our method super-resolved the characters accurately.

TABLE III
RECOGNITION RATES ACHIEVED ON REAL-WORLD IMAGES.

Test Images	# Correct Characters		
	All	≥ 6	≥ 5
HR (original images)	90.6%	98.7%	100%
LR (degraded images)	9.9%	28.0%	56.2%
LR + SR (SR3 [35])	31.7%	63.7%	80.1%
LR + SR (PLNET [9])	36.3%	67.2%	82.5%
LR + SR (**Proposed**)	**39.5%**	**70.2%**	**83.1%**

V. CONCLUSIONS

This article proposes a specialized super-resolution method designed to improve the readability of characters and enhance recognition rates in LPR applications. Our approach involves the implementation of the Layout and Character Oriented Focal Loss (LCOFL) to guide the network in accurately reconstructing characters according to the LP layout, effectively mitigating confusion between structurally similar characters. Additionally, we enhanced the PLTFAM model [9]

Fig. 4. Super-resolved LPs generated by our method and baselines from real-world images. The background image shows the original scene from which the LR image was extracted. From top to bottom: LP reconstructions by SR3 [35], PLNET [9], our method, and a reference HR image from a different frame.

by introducing shared weights and integrating deformable convolutions, leading to improved feature extraction.

Our experiments were conducted on the diverse RodoSol-ALPR [33] dataset. The results revealed significantly improved recognition rates in images reconstructed using the proposed method compared to state-of-the-art approaches. Remarkably, our method led to a recognition rate of 49.8% being achieved by the OCR model, whereas the methods proposed in [35] and [9] led to recognition rates of 43.1% and 39.0%, respectively.

The dataset and source code for all experiments are publicly available at https://github.com/valfride/lpsr-lacd.

While our experiments were limited to Brazilian and Mercosur LPs, the findings provide a foundation for future research to explore the method's efficacy across different LP layouts.

ACKNOWLEDGMENTS

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)* - Finance Code 001, and in part by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)* (# 315409/2023-1 and # 312565/2023-2). We thank the support of NVIDIA Corporation with the donation of the Quadro RTX 8000 GPU 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] D. Moussa *et al.*, "Forensic license plate recognition with compression-informed transformers," in *IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 406–410.

[3] 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.

[4] 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.

[5] 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.

[6] Z. Rao *et al.*, "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.

[7] M. Santos *et al.*, "Face super-resolution using stochastic differential equations," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2022, pp. 216–221.

[8] V. Nascimento *et al.*, "Combining attention module and pixel shuffle for license plate super-resolution," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2022, pp. 228–233.

[9] ——, "Super-resolution of license plate images using attention modules and sub-pixel convolution layers," *Computers & Graphics*, vol. 113, pp. 69–76, 2023.

[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] A. Mehri, P. B. Ardakani, and A. D. Sappa, "MPRNet: Multi-path residual network for lightweight image super resolution," in *IEEE Winter Conference on Applications of Computer Vision*, 2021, pp. 2703–2712.

[12] A. Liu, Y. Liu, J. Gu, Y. Qiao, and C. Dong, "Blind image super-resolution: A survey and beyond," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5461–5480, 2023.

[13] S. M. Silva and C. R. Jung, "Real-time license plate detection and recognition using deep convolutional neural networks," *Journal of Visual Communication and Image Representation*, p. 102773, 2020.

[14] I. O. Oliveira *et al.*, "Vehicle-Rear: A new dataset to explore feature fusion for vehicle identification using convolutional neural networks," *IEEE Access*, vol. 9, pp. 101 065–101 077, 2021.

[15] X. Ke *et al.*, "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.

[16] Q. Liu *et al.*, "Improving multi-type license plate recognition via learning globally and contrastively," *IEEE Transactions on Intelligent Transportation Systems*, pp. 1–11, 2024, early Access.

[17] G. R. Gonçalves *et al.*, "Real-time automatic license plate recognition through deep multi-task networks," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2018, pp. 110–117.

[18] X. Fan and W. Zhao, "Improving robustness of license plates automatic recognition in natural scenes," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 10, pp. 18 845–18 854, 2022.

[19] Y. Wang *et al.*, "Rethinking and designing a high-performing automatic license plate recognition approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 8868–8880, 2022.

[20] A. Maier *et al.*, "Reliability scoring for the recognition of degraded license plates," in *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2022, pp. 1–8.

[21] M. Lin, L. Liu, F. Wang, J. Li, and J. Pan, "License plate image reconstruction based on generative adversarial networks," *Remote Sensing*, vol. 13, no. 15, p. 3018, 2021.

[22] C. Ledig *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 105–114.

[23] A. Hamdi *et al.*, "A new image enhancement and super resolution technique for license plate recognition," *Heliyon*, vol. 7, no. 11, 2021.

[24] S. Pan, S.-B. Chen, and B. Luo, "A super-resolution-based license plate recognition method for remote surveillance," *Journal of Visual Communication and Image Representation*, vol. 94, p. 103844, 2023.

[25] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *European Conference on Computer Vision (ECCV)*, 2019, pp. 63–79.

[26] D. Kim *et al.*, "AFA-Net: Adaptive feature attention network in image deblurring and super-resolution for improving license plate recognition," *Computer Vision and Image Understanding*, vol. 238, p. 103879, 2024.

[27] S. Lee *et al.*, "Super-resolution of license plate images via character-based perceptual loss," in *IEEE International Conference on Big Data and Smart Computing (BigComp)*, 2020, pp. 560–563.

[28] B. Lorch, S. Agarwal, and H. Farid, "Forensic reconstruction of severely degraded license plates," *Electronic Imaging*, vol. 31, pp. 1–7, 2019.

[29] S. Jamil, "Review of image quality assessment methods for compressed images," *Journal of Imaging*, vol. 10, no. 5, p. 113, 2024.

[30] M. Cheon *et al.*, "Ambiguity of objective image quality metrics: A new methodology for performance evaluation," *Signal Processing: Image Communication*, vol. 93, p. 116150, 2021.

[31] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[32] X. Zhu, H. Hu, S. Lin, and J. Dai, "Deformable ConvNets V2: More deformable, better results," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 9300–9308.

[33] 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.

[34] ——, "A first look at dataset bias in license plate recognition," in *Conference on Graphics, Patterns and Images*, Oct 2022, pp. 234–239.

[35] C. Saharia *et al.*, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4713–4726, 2023.