This is the author-prepared version of a paper accepted for publication in the *Journal of the Brazilian Computer Society*.
The final version is available in the SBC OpenLib Library (DOI: 10.5753/jbcs.2025.5159).

Toward Advancing License Plate Super-Resolution in Real-World Scenarios: A Dataset and Benchmark

Valfride Nascimento [**Federal University of Paraná** | ***vwnascimento@inf.ufpr.br***]
Gabriel E. Lima [**Federal University of Paraná** | ***gelima@inf.ufpr.br***]
Rafael O. Ribeiro [**Brazilian Federal Police** | ***rafael.ror@pf.gov.br***]
William Robson Schwartz [**Federal University of Minas Gerais** | ***william@dcc.ufmg.br***]
Rayson Laroca [**Pontifical Catholic University of Paraná, Federal University of Paraná** | ***rayson@ppgia.pucpr.br***]
David Menotti [**Federal University of Paraná** | ***menotti@inf.ufpr.br***]

Department of Informatics, Federal University of Paraná, R. Evaristo F. Ferreira da Costa 391, Jardim das Américas, Curitiba, PR, 81530-090, Brazil.

Abstract. Recent advancements in super-resolution for License Plate Recognition (LPR) have sought to address challenges posed by low-resolution (LR) and degraded images in surveillance, traffic monitoring, and forensic applications. However, existing studies have relied on private datasets and simplistic degradation models. To address this gap, we introduce UFPR-SR-Plates, a novel dataset containing 10,000 tracks with 100,000 paired low and high-resolution license plate images captured under real-world conditions. We establish a benchmark using multiple sequential LR and high-resolution (HR) images per vehicle – five of each – and two state-of-the-art models for super-resolution of license plates. We also investigate three fusion strategies to evaluate how combining predictions from a leading Optical Character Recognition (OCR) model for multiple super-resolved license plates enhances overall performance. Our findings demonstrate that super-resolution significantly boosts LPR performance, with further improvements observed when applying majority vote-based fusion techniques. Specifically, the Layout-Aware and Character-Driven Network (LCDNet) model combined with the Majority Vote by Character Position (MVCP) strategy led to the highest recognition rates, increasing from 1.7% with low-resolution images to 31.1% with super-resolution, and up to 44.7% when combining OCR outputs from five super-resolved images. These findings underscore the critical role of super-resolution and temporal information in enhancing LPR accuracy under real-world, adverse conditions. The proposed dataset is publicly available to support further research and can be accessed at: `https://valfride.github.io/nascimento2025toward/`.

Keywords: Ensemble, License Plate Recognition, Super-Resolution, UFPR-SR-Plates dataset.

1 Introduction

License Plate Recognition (LPR) systems have become increasingly popular across various practical applications, such as traffic monitoring and toll collection [Laroca *et al.*, 2021; Ke *et al.*, 2023; Liu *et al.*, 2024b]. These systems are designed to accurately recognize characters on a License Plate (LP) after it has been detected within an image.

While recent studies on LPR have reported high recognition rates, the results are primarily based on experiments using high-resolution (HR) images, where the LP characters are clearly defined and free from significant noise [Silva and Jung, 2022; Laroca *et al.*, 2023b; Rao *et al.*, 2024]. However, accurately recognizing characters in low-resolution (LR) or degraded images remains a significant challenge.

In surveillance scenarios, images are often captured at low resolutions or are heavily compressed due to constraints in storage and bandwidth. Hence, LP characters may become distorted, blend into the background, or overlap with neighboring characters, making recognition challenging. This underscores the need for robust methods that can effectively handle these types of degradation.

Taking this into account, various image enhancement techniques, including super-resolution (SR), have been proposed to improve image quality [Moussa *et al.*, 2022; Nascimento *et al.*, 2023, 2024a; Pan *et al.*, 2024]. Although these techniques aim to improve LPR, their performance is frequently assessed using metrics such as Structural Similarity Index Measure (SSIM) and Peak Signal-to-Noise Ratio (PSNR), which are known not to correlate well with human assessment of visual quality or recognition accuracy [Johnson *et al.*, 2016; Zhang *et al.*, 2018; Mehri *et al.*, 2021; Liu *et al.*, 2023]. Furthermore, most studies relied solely on private datasets [Hamdi *et al.*, 2021; Maier *et al.*, 2022; Luo *et al.*, 2024], hindering fair comparisons. Many also adopted simplistic degradation models, where LR images were created by simply downsampling the original HR images [Nascimento *et al.*, 2022; Kim *et al.*, 2024; Pan *et al.*, 2024].

In response to these limitations, we introduce a new publicly available dataset, UFPR-SR-Plates[1]. It comprises 100,000 LP images captured by a rolling shutter camera installed on a Brazilian road. UFPR-SR-Plates includes 10,000 LP tracks, each consisting of ten consecutive images – five LR images captured when the vehicle was farthest from the camera, and five HR images captured at its closest point. These tracks, recorded under varying environmental and lighting conditions, feature two distinct LP layouts: Brazilian and Mercosur. Unlike synthetically generated datasets,

[1]The UFPR-SR-Plates dataset is publicly available at `https://valfride.github.io/nascimento2025toward/`

UFPR-SR-Plates offers a more accurate representation of surveillance scenarios and makes it a valuable resource for advancing LP super-resolution research.

In summary, the main contributions of this work are:

- A publicly available dataset containing 100,000 images, divided into 10,000 tracks, with each track containing five LR images and five HR images of the same LP. To enhance variability, 5,000 tracks were collected at a resolution of 1280×960 pixels, while the remaining 5,000 tracks were captured at 1920×1080 pixels. The dataset is evenly distributed between Mercosur and Brazilian LPs, making it the largest dataset in terms of the number of LPs for both layouts;
- We conducted benchmark experiments on the proposed dataset using five state-of-the-art super-resolution models: (i) general-purpose approaches (SR3 [Saharia *et al*., 2023], Real-ESRGAN [Wang *et al*., 2021]), and (ii) LP-specialized networks (LPSRGAN [Pan *et al*., 2024], PLNET [Nascimento *et al*., 2024a], and LCD-Net [Nascimento *et al*., 2024b]). For each track, we generated five super-resolved images from the LR images and compared the recognition results obtained by the leading Optical Character Recognition (OCR) model, GP_LPR [Liu *et al*., 2024b]. The super-resolution process significantly boosted recognition accuracy, increasing from 2.2% to 29.9% for a single super-resolved image. To further enhance LPR performance, we explored three fusion strategies for combining the outputs from the OCR model based on multiple super-resolved images. Notably, applying the Majority Vote by Character Position (MVCP) strategy with five super-resolved images improved the recognition rate from 29.9% to 42.3%. The proposed dataset enables the exploration of temporal relationships among low-resolution LPs, as it includes multiple sequential LR images for each LP.

The remainder of this paper is structured as follows. Section 2 provides a brief review of related works. In Section 3, we introduce the UFPR-SR-Plates dataset. The experiments are detailed in Section 4. Finally, Section 5 concludes the paper by summarizing our findings and their significance.

2 Related Work

This section provides an overview of relevant works in LPR and LP super-resolution. More specifically, Section 2.1 covers recent advancements and techniques in LPR, highlighting key innovations and their impact on recognition accuracy. Section 2.2 discusses the integration of super-resolution methods with LPR, focusing on their role in improving the recognition of low-quality or degraded LP images.

2.1 License Plate Recognition (LPR)

The primary goal of LPR is to accurately identify characters from a given LP image. To tackle this challenge, Silva and Jung [2020] proposed treating the LPR stage as an object detection task, where each character class is identified as a distinct object. They introduced CR-NET, a model based on YOLO [Redmon *et al*., 2016], which has shown significant effectiveness for LPR in subsequent studies [Laroca *et al*., 2021; Oliveira *et al*., 2021; Silva and Jung, 2022].

Recently, advancements in the field have moved toward holistic treatment of the entire LP image for text recognition, departing from traditional segmentation methods that isolate individual characters. This shift has improved recognition accuracy while also enhancing computational efficiency. For instance, Ke *et al*. [2023] introduced a lightweight, multi-scale LPR network that integrates global channel attention layers to effectively fuse low- and high-level features.

To address real-world challenges, such as substantially tilted LPs caused by suboptimal camera positioning, recent research has focused on incorporating attention mechanisms into deep learning models. Rao *et al*. [2024] integrated attention mechanisms into a CRNN model, while Liu *et al*. [2024a] introduced deformable spatial attention modules to enhance feature extraction and capture the LP's global layout. Building on this, Liu *et al*. [2024b] presented a robust OCR model called Global Perception License Plate Recognition (GP_LPR) for recognizing irregular LPs through the use of deformable spatial attention and global perception modules (this model is further detailed in Section 4.1). These advancements collectively address challenges such as attention deviation and character misidentification, leading to state-of-the-art performance on popular datasets such as CCPD [Xu *et al*., 2018] and RodoSol-ALPR [Laroca *et al*., 2022].

Although these models have reported impressive accuracy and inference speed, most evaluations were conducted on datasets where all LP characters are clearly legible, even on challenging scenarios involving tilted LPs. This setup, however, does not accurately reflect real-world surveillance environments, where cost-effective cameras are typically used and bandwidth limitations often degrade image quality. Consequently, LR images with blurry characters that blend into adjacent ones and the LP background are prevalent, posing a substantial challenge for robust LPR [Moussa *et al*., 2022; Ke *et al*., 2023; Schirrmacher *et al*., 2023].

2.2 Super-Resolution for LPR

The quality of an image is affected by various factors, including lighting, weather, camera distance, motion blur, and storage techniques, each introducing unique noise patterns. These factors, combined with the structural variability in low-resolution LPs, make it challenging for LPR systems to accurately identify characters in such images. Although recent advancements in SR methods have shown potential in improving character visibility in low-quality images [Liu *et al*., 2023], the specific challenges related to LPR under degraded conditions remain largely unaddressed [Maier *et al*., 2022; Hijji *et al*., 2023; Angelika Mulia *et al*., 2024].

To address these challenges, Lin *et al*. [2021] proposed an ESRGAN-based [Wang *et al*., 2019] approach for LP image enhancement called PatchGAN. This method leverages a residual dense network with progressive upsampling to preserve high-frequency details effectively. While PatchGAN achieved impressive PSNR and SSIM values, its performance

gains over other Generative Adversarial Network (GAN)-based methods (e.g., SRGAN [Ledig *et al*., 2017]) were relatively modest, and the model's complexity may limit its applicability in real-time scenarios.

Expanding on this, Hamdi *et al*. [2021] proposed an approach called Double Generative Adversarial Networks for Image Enhancement and Super-Resolution, which employs two sequential networks: the first to deblur the image and the second to apply super-resolution, yielding the final output. While their approach demonstrated effectiveness, it was only tested on synthetically generated low-resolution images created from high-resolution LPs.

Recognizing the gap in integrating character recognition into the SR process, Lee *et al*. [2020] introduced a perceptual loss based on features extracted from scene text recognition models. Specifically, they leveraged intermediate representations from ASTER [Shi *et al*., 2019] to train a model based on GANs. Their experiments showed that adding this perceptual loss improved results compared to models trained without it. However, a lack of detailed information on datasets and degradation methods hinders the reproducibility and generalization of their approach.

Building on these developments, Pan *et al*. [2023] introduced a complete pipeline for the SR of LPs followed by recognition, utilizing ESRGAN [Wang *et al*., 2019] for single-character enhancement. Their method demonstrated effectiveness with moderately low-quality LPs, nevertheless, it struggled with severely degraded images, particularly when character boundaries were unclear. To overcome these challenges, Pan *et al*. [2024] developed LPSRGAN, which processes the entire LP image and incorporates a degradation model that generates more realistic low-resolution LPs. Despite these improvements, LPSRGAN still struggled to reconstruct characters under severely degraded conditions.

Kim *et al*. [2024] introduced AFA-Net, an architecture that integrates deblurring sub-networks at both the pixel and feature levels. Their method was evaluated on a dataset containing low-resolution and blurred LP images captured from unconstrained dash cams. While the results were promising, the LR images were artificially generated through simple interpolation, and the testing protocol focused solely on super-resolving digits, excluding letter recognition. This limits the generalizability of their approach to real-world LP images.

Luo *et al*. [2024] designed a domain-specific degradation model to simulate real-world LP degradations, incorporating common factors such as motion blur, lighting issues, and noise. By retraining ESRGAN on a dataset of high-resolution LPs with these simulated degradations, they achieved robust recognition performance. This success highlights the model's effectiveness in controlled settings where degradation types are aligned with the training data. However, when confronted with unconstrained real-world scenarios, with severe occlusion, extreme lighting variations, or unforeseen blur, the method struggled to preserve accurate character structure, indicating that further refinement is necessary for application in complex, real-world settings.

AlHalawani *et al*. [2024] recently introduced DiffPlate, a diffusion model for LP super-resolution that outperformed ESRGAN and SwinIR [Liang *et al*., 2021] in terms of PSNR and SSIM. However, its high computational cost restricts its real-time applicability in surveillance systems. Furthermore, the model was trained and tested using synthetically generated images, where high-resolution LPs were downsampled by a factor of four to produce low-resolution counterparts.

Nascimento *et al*. [2023, 2024a] introduced the Pixel-Level Network (PLNET), a super-resolution model that employs specialized attention modules to enhance the quality of LP images. While PLNET showed potential in improving character clarity in LR scenarios, the experiments were limited to synthetically generated LP images. In subsequent research, the same authors [Nascimento *et al*., 2024b] developed the Layout-Aware and Character-Driven Network (LCDNet), which further improved character structure and positioning in LP layouts through deformable convolutions, shared-weight attention modules, and a GAN-based approach with an OCR discriminator and a layout-aware perceptual loss. Although this latter study included preliminary experiments with real-world images, the dataset was not made publicly available, which limits reproducibility. Both PLNET and LCDNet are described in more detail in Section 4.1, as they are utilized in our experiments.

Sendjasni and Larabi [2024] proposed RDASRNet, a framework designed for extreme license plate SR (with a $\times 16$ scaling factor). The architecture integrates a hierarchical channel attention mechanism that iteratively refines features extracted from LR inputs. To enhance training, a dual-loss strategy was employed – combining mean squared error with a contrastive loss guided by a Siamese network. This approach enforces perceptual and structural consistency between the super-resolved outputs and HR ground-truth images within a latent space. While RDASRNet achieves state-of-the-art performance on synthetic benchmarks, its high computational cost poses challenges for real-time deployment in surveillance systems. Furthermore, its reliance on synthetic training data – where LR images are generated via idealized downsampling – raises concerns about its generalizability to real-world degradations, a limitation shared with other studies [Pan *et al*., 2023; AlHalawani *et al*., 2024].

In summary, while substantial progress has been made in super-resolution techniques for LPR, a major barrier to further advancement is the limited availability of public datasets containing paired low- and high-resolution LPs. Most existing research relies on either proprietary datasets [Lee *et al*., 2020; Hamdi *et al*., 2021; Maier *et al*., 2022; Pan *et al*., 2024] or synthetically generated LR images [Pan *et al*., 2023; AlHalawani *et al*., 2024; Kim *et al*., 2024; Luo *et al*., 2024; Sendjasni and Larabi, 2024].

3 The UFPR-SR-Plates Dataset

The UFPR-SR-Plates dataset comprises 10,000 tracks, each with five consecutive LR images and five consecutive HR images of the same LP. With a total of 100,000 images, this dataset is well-suited for SR and LPR tasks. It offers real-world images captured under diverse noise and degradation conditions, while enabling direct LR-HR comparison and analysis of temporal variations within frame sequences.

The images were taken with a rolling shutter camera near

Figure 1. Examples of scenarios from which the LP images in the UFPR-SR-Plates dataset were extracted. These images showcase a variety of vehicle types and their corresponding LPs, captured under different environmental conditions. The first row shows images taken with a resolution of 1280 × 960 pixels, while the second row displays images captured at 1920 × 1080 pixels. For better visualization, all images in this figure were slightly resized.

Figure 2. Examples of tracks from the UFPR-SR-Plates dataset. Each track comprises five consecutive LR images and five consecutive HR images of the same LP, captured under varying conditions. Each row shows a single track, with the LR images displayed on the left and the corresponding HR images on the right. We remark that even what we consider 'HR' in the context of this work is of lower quality than the datasets typically used in LPR research.

the Department of Informatics at the Federal University of Paraná in Curitiba, Brazil, simulating real-world surveillance conditions. Figure 1 presents sample images collected to build the UFPR-SR-Plates dataset (i.e., prior to LP detection and extraction), highlighting the diversity of vehicle types, including cars, buses, and trucks.

Data collection occurred over six months, with images captured daily during a fixed 10-hour interval, from 7 a.m. to 5 p.m. Nighttime images were excluded due to infrared interference, which often caused overexposure or underexposure, making LP recognition challenging even in HR images.

The original images were evenly distributed between two video resolutions: 1280 × 960 pixels and 1920 × 1080 pixels. This variation in resolution enables the extraction of LPs with different pixel densities, allowing for a broader evaluation of SR methods across varying image quality. Consequently, UFPR-SR-Plates becomes more versatile for tasks that require adaptation to different levels of detail.

The proposed dataset is also balanced between two LP layouts: Brazilian and Mercosur. Brazilian LPs follows a format of three letters followed by four digits, while Mercosur LPs features three letters, one digit, one letter, and two digits. Although both types of LPs are similar in size and shape, they differ significantly in color schemes and character fonts.

As shown in Figure 3, vehicle tracks were derived from video footage capturing vehicles entering and exiting the road from opposite sides. To detect and track vehicles, we employed the YOLOv8 model [Ultralytics, 2023], a choice motivated by the proven effectiveness of the YOLO family in object detection in unconstrained scenarios [Lima *et al*., 2024; Laroca *et al*., 2021, 2025]. We fine-tuned the model to meet our specific needs, starting with a pre-trained version and collecting initial bounding box annotations for detected vehicles. After each detection round, we carefully reviewed and manually corrected any annotation errors, incorporating these adjustments into the training set and retraining the model. This iterative process progressively improved detection accuracy, culminating in a refined dataset of 2,954 labeled images tailored to our application.

We cropped each vehicle's region of interest using the annotations from the process described above. Subsequently, we applied IWPOD-NET [Silva and Jung, 2022] to locate the LP corners. Although IWPOD-NET is a well-regarded method for this task [Jia and Xie, 2023; Wei *et al*., 2024], its original training on high-quality images limited its effectiveness in our dataset, particularly for vehicles at greater distances. To overcome this limitation, we retrained IWPOD-NET from scratch with optimized hyperparameters, significantly improving its robustness in detecting LPs from distant vehicles. More specifically, we started with an initial set of 300 annotated LPs to train the IWPOD-NET model. Through an iterative process, we tested the model on new images, corrected any errors, and progressively expanded the training set with these refined annotations. After several iterations, we conducted a final training phase with 839 images, ensuring precise LP corner detection considering our scenario. We then applied the model to extract LPs using a minimum bounding box method, adding 20% padding to both vertical and horizontal dimensions to capture contextual surroundings.

From the resulting sequences, we selected the five LR images that were farthest from the camera. To annotate the LP

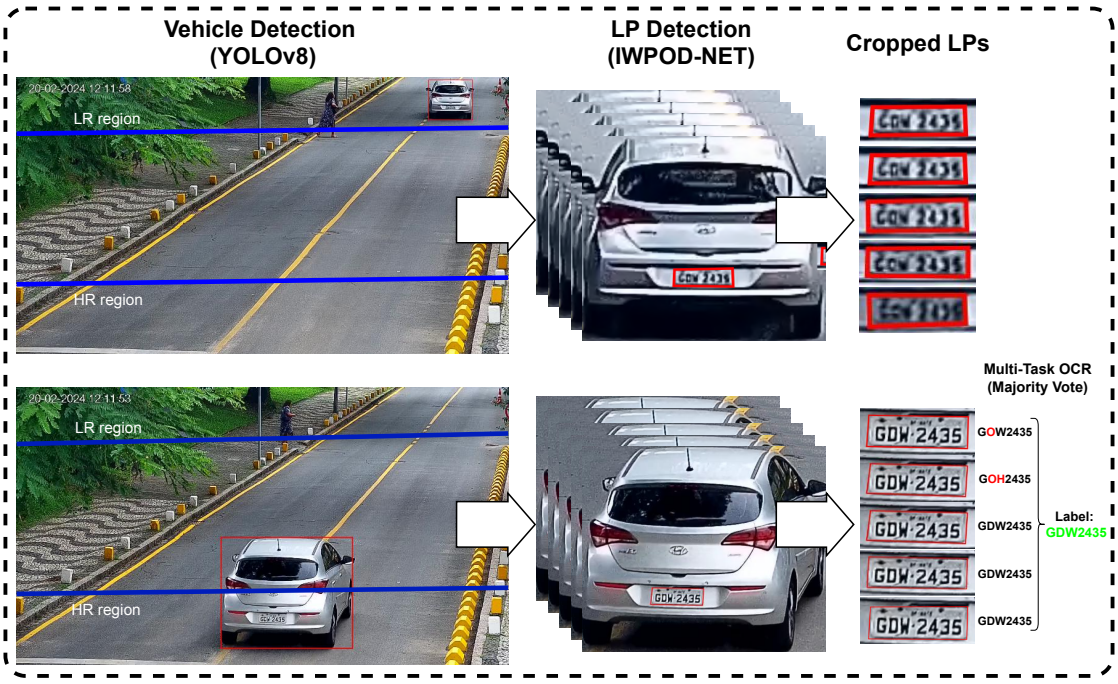

Figure 3. Illustration of the process of vehicle identification and tracking using *YOLOv8* [Ultralytics, 2023], followed by LP corner detection with IWPOD-NET [Silva and Jung, 2022]. LP images were extracted from the original frames based on the detected corners. A multi-task OCR model, proposed by Gonçalves *et al*. [2018], was applied to recognize the text on the extracted LP images. Majority voting was applied to determine the final label. The blue lines in the original images demarcate the regions for detecting low-resolution LPs (above the higher line) and high-resolution LPs (below the lower line).

characters, we employed the multi-task OCR developed by Gonçalves *et al*. [2018] on each of the five HR samples in the track, utilizing a sequence-level majority vote strategy. Figure 2 shows all LP images from five tracks in the dataset.

Each image within a track is also accompanied by a JSON file containing the coordinates $(x,\ y)$ of its four corners, the layout of the LP (Brazilian or Mercosur), and its textual content (e.g., ABC-1234). Table 1 presents a summary of key statistical characteristics for each layout across both resolutions in the UFPR-SR-Plates dataset, including the median, maximum, and minimum dimensions, as well as the total number of unique LPs.

Table 1. Summary of statistics (in pixels) for Brazilian and Mercosur LPs across resolutions in the UFPR-SR-Plates dataset.

UFPR-SR-Plates	1280 × 960				1920 × 1080			
	Brazilian		Mercosur		Brazilian		Mercosur	
	LR	HR	LR	HR	LR	HR	LR	HR
Median Height	19	34	18	32	21	50	21	38
Median Width	35	69	35	67	49	100	49	100
Max Height	28	60	25	49	28	52	26	52
Min Height	15	22	13	23	14	26	16	26
Max Width	56	103	59	88	35	122	61	122
Min Width	26	51	24	50	34	71	34	71
Unique LPs	1,663		2,500		1,627		2,496	

Although the dataset acquisition process was automated, all annotations were manually reviewed to ensure the reliability of the UFPR-SR-Plates dataset for research purposes. Despite the popularity of the chosen OCR model in the literature [Gonçalves *et al*., 2019; Nascimento *et al*., 2024b], we found that it produced errors in approximately 5% of the tracks. These errors were rectified during the aforementioned analysis process.

For the experimental protocol, we divided each resolution set (1920 × 1080 and 1280 × 960 pixels) into approximately 40%, 20%, and 40% for training, validation, and testing, respectively. This resulted in 3,965 tracks for training, 2,030 for validation, and 4,005 for testing. Note that the number of tracks in the training, validation, and test sets does not conform to the exact ratios of 4,000/2,000/4,000 images, as we carefully avoided any overlap of LPs across the training, validation, and test sets, even when the same vehicle/LP was depicted in images of different resolutions. For instance, if the LP "ABC-1234" is included in the training set, it is strictly excluded from both the test and validation sets. As demonstrated by Laroca *et al*. [2023a], the presence of near-duplicate LP images across different subsets can artificially inflate model performance. By preventing such overlaps, UFPR-SR-Plates creates a fair and reliable resource for training, validating, and testing deep learning-based models.

Regarding privacy concerns, LPs of vehicles registered in Brazil are not associated with the personal information of the vehicle owner, thus mitigating the risk of privacy breaches. Each LP uniquely identifies the vehicle itself [Presidência da República, 2014; Oliveira *et al*., 2021].

4 Experimental Results

This section delves into the experimental details of this work. Section 4.1 introduces the models employed, providing an overview of the framework and key hyperparameters. Section 4.2 outlines the fusion strategies implemented to combine the predictions generated by the OCR model for multiple super-resolved images, aiming to enhance LPR performance. Finally, Section 4.3 presents and discusses the results.

All experiments were conducted on a computer equipped with an AMD Ryzen 5950X 3.4 GHz CPU, 128 GB of RAM, an SSD with read speeds of 535 MB/s and write speeds of

445 MB/s, and an NVIDIA Quadro RTX 8000 GPU (48 GB).

4.1 Models

We applied the GP_LPR model [Liu *et al*., 2024b] to LPR on super-resolved images generated from the LR images in the UFPR-SR-Plates dataset. The SR process was conducted using five state-of-the-art networks: (i) two general-purpose models: Real-ESRGAN [Wang *et al*., 2021] and SR3 [Saharia *et al*., 2023], and (ii) three LP-specialized models LPSRGAN [Pan *et al*., 2024], PLNET [Nascimento *et al*., 2023, 2024a], and LCDNet [Nascimento *et al*., 2024b].

The chosen models achieved state-of-the-art performance in both general SR tasks [Saharia *et al*., 2023; Luo *et al*., 2024] and LP-specific SR tasks [Nascimento *et al*., 2023; Pan *et al*., 2024]. Due to the absence of an official implementation, we reimplemented LPSRGAN based on the methodology described by Pan *et al*. [2024]. The code for all models is publicly available[2,3,4,5,6].

The GP_LPR model focuses on LPR of irregular LPs, employing attention mechanisms to handle perspective distortion. Central to its architecture is the global perception module, which enhances character feature completeness by incorporating global visual information. This facilitates global interaction among features, distinguishing characters with similar structures and minimizing misidentifications. Additionally, the model employs the Deformable Spatial Attention Module, featuring deformable convolution layers that adjust to variations in character positions and shapes, improving the network's ability to capture the overall layout of LPs.

Real-ESRGAN [Wang *et al*., 2021] extends ESRGAN [Wang *et al*., 2018] by introducing a high-order degradation model tailored for real-world scenarios. Unlike traditional approaches, it trains exclusively on synthetically degraded data generated through a rigorous pipeline that applies multiple degradation steps – including blurring, noise injection, resizing, and compression. A key innovation lies in its artifact suppression mechanism, which employs filtering to mitigate distortions introduced during degradation simulation. The framework adopts a U-Net discriminator with spectral normalization to stabilize adversarial training. Training proceeds in two stages: first, a PSNR-oriented optimization ensures structural fidelity, followed by a perceptual refinement stage to enhance visual quality. This methodology has demonstrated state-of-the-art performance in general image restoration, validating the effectiveness of synthetic degradation modeling for practical super-resolution tasks.

SR3 [Saharia *et al*., 2023] takes a distinct approach by framing super-resolution as a diffusion process. Unlike common GAN-based methods, Super-Resolution via Iterative Refinement (SR3) iteratively refines a noisy input image into a high-resolution output through a Markov chain. This stochastic refinement enables the model to explore multiple plausible reconstructions, particularly advantageous for recovering fine-grained details in severely degraded LP images. The iterative process is guided by a noise prediction network trained to reverse a predefined degradation schedule, making SR3 robust to diverse noise types and compression artifacts.

LPSRGAN [Pan *et al*., 2024] enhances LPR accuracy in unconstrained scenarios through a three-pronged approach. It introduces an n-stage random combination degradation (n-RCD) model to simulate real-world degradations like blur, noise, and compression via multi-stage randomized combinations, addressing limitations of simplistic degradation pipelines. The framework adopts a modified RRDBNet+ generator, building on Residual-in-Residual Dense Blocks (RRDB) [Wang *et al*., 2019] with dropout layers to improve feature representation and generalization across LP layouts. To prioritize character clarity for recognition systems, LPSRGAN employs a perceptual loss optimized for LPR, aligning super-resolved images with a Connectionist Temporal Classification loss to optimize character clarity by aligning super-resolved images with OCR output predictions. This integration of realistic degradation modeling, architectural enhancements, and task-specific optimization enables robust restoration of degraded LPs, particularly under severe real-world distortions.

The PLNET model builds upon the foundation laid by Mehri *et al*. [2021], introducing refinements specifically tailored for LP super-resolution. It incorporates a shallow feature extractor module, using an autoencoder equipped with *PixelShuffle* and *PixelUnshuffle* layers [Shi *et al*., 2016] to efficiently extract shallow features while preserving essential information through skip connections. PLNET also integrates a mechanism to capture inter-channel and spatial relationships, thus improving the model's ability to rearrange and scale input data more effectively than conventional methods.

LCDNet employs deformable convolutions and shared-weight attention modules within a GAN framework. An OCR model acts as the discriminator, steering the super-resolution process toward prioritizing LPR accuracy. During training, LCDNet optimizes a loss function designed to preserve character structure and the overall integrity of the LP layout (e.g., penalizing any confusion between letters and digits). This ensures both visually clear and accurate LP recognition.

Here we detail the key hyperparameters used for training the models, all implemented using the PyTorch framework. The hyperparameters were selected based on the prior works [Saharia *et al*., 2023; Pan *et al*., 2024; Wang *et al*., 2021; Nascimento *et al*., 2024a,b; Liu *et al*., 2024b] and preliminary experiments conducted on the validation set of the proposed dataset. For all models, we employed the Adam optimizer. In LPSRGAN [Pan *et al*., 2024], we replaced the original HyperLPR3 OCR with Gonçalves *et al*. [2019]'s multi-task model trained on UFPR-SR-Plates (ensuring fair comparison as HyperLPR3 was trained on Chinese plates with unavailable source code), using generator/discriminator learning rates of $10^{-4}/10^{-5}$ respectively. Real-ESRGAN [Wang *et al*., 2021] followed its standard two-stage protocol: PSNR-oriented pretraining for 10^6 iterations ($lr = 2{\times}10^{-4}$) followed by GAN fine-tuning for $4{\times}10^5$ iterations ($lr = 1{\times}10^{-4}$), with EMA stabilization and U-Net discriminator. SR3 [Saharia *et al*., 2023] employed 100 denoising steps across 10^6 training iterations ($lr = 1{\times}10^{-4}$),

[2]GP_LPR: https://github.com/mmm2024/gp_lpr/
[3]PLNET: https://github.com/valfride/lpr-rsr-ext/
[4]LCDNet: https://github.com/valfride/lpsr-lacd/
[5]Real-ESRGAN: https://github.com/xinntao/Real-ESRGAN/
[6]LPSRGAN: https://github.com/valfride/lpsrgan/

incorporating Gaussian blur augmentation on low-resolution inputs. For both LCDNet and PLNET, the initial learning rate was set to 10^{-4}, decaying by a factor of 0.8 when no improvement in the loss function was observed. For the GP_LPR model, the maximum decoding length was set to $K = 7$ to match the 7-character format of the LPs in the UFPR-SR-Plates dataset. To mitigate training oscillations in GP_LPR, following Liu *et al*. [2024b], we applied a step learning rate scheduler starting with an initial learning rate of 10^{-3} and decaying by a factor of 0.8 every 50 epochs.

Additionally, we added padding with gray pixels to both the LR and HR images to preserve their aspect ratio before resizing them to dimensions of 16×48 and 32×96 pixels, respectively, corresponding to an upscale factor of 2.

4.2 Fusion Methods

Inspired by the work of Laroca *et al*. [2023b], we examine three fusion methods to combine predictions from multiple super-resolved images generated by the SR networks. For each track in the test set, we process N consecutive LR images of the same LP, captured at varying distances from the camera. Each LR image is independently super-resolved and fed to the OCR model. The predictions are then fused using one of the following strategies. In our experiments, $N \in \{1, 3, 5\}$, with images selected sequentially from nearest to farthest. This process is illustrated in Figure 4.

The first fusion method, Highest Confidence (HC), straightforwardly selects the single prediction with the highest associated confidence score as the final output:

$$\hat{y} = y_k, \text{ where } k = \arg \max_{i \in \{1,\ldots,N\}} P_i,$$

where P_i is the confidence score for the i-th OCR prediction y_i. This strategy aligns with classical confidence-based fusion rules (e.g., [Kittler *et al*., 1998]), which prioritize predictions with the highest certainty.

The second method, Majority Vote (MV), is an ensemble learning technique [Zhou, 2025] that selects the most frequent prediction among all N outputs:

$$\hat{y} = \text{mode}\left(\{y_1, y_2, \ldots, y_N\}\right),$$

for mode defined as:

$$\text{mode}(S) = \arg \max_{x \in S} \text{count}_S(x),$$

where $\text{count}_S(x)$ is the number of times x appears in the multiset S. MV has been widely adopted in applications such as handwriting recognition [Zhao and Liu, 2020].

The third method, Majority Vote by Character Position (MVCP), aggregates predictions per character position:

$$\hat{y}_j = \text{mode}\left(\{y_{(1,j)}, y_{(2,j)}, \ldots, y_{(N,j)}\}\right),$$

where $y_{(i,j)}$ denotes the j-th character (for $j \in [1, 7]$, $j \in \mathbb{N}$) of the i-th OCR prediction. The final output $\hat{y}$ is formed by concatenating all $\hat{y}_j$. This hierarchical approach draws inspiration from structured prediction frameworks [Ghamrawi and McCallum, 2005], resolving ambiguities at each character position.

A key challenge in majority-vote strategies is resolving ties between competing predictions. To address this, our approach prioritizes the prediction with the highest average character-level confidence score within tied groups. Consider five OCR predictions for a Brazilian LP with associated confidence scores: two instances of "ABC-1234" (0.95 and 0.91), two instances of "ABD-1234" (0.89 and 0.88), and "HBG-1284" (0.71). Here, the MV method resolves the tie between the top two candidates ("ABC-1234" and "ABD-1234") by comparing their confidence scores – 0.93 for "ABC-1234" and 0.88 for "ABD-1234" – selecting the former. For character-level ambiguities (e.g., a conflict between two "C"s and two "D"s in the third position), the MVCP strategy defaults to the character in the prediction with the highest confidence (in this case selecting "C" from "ABC-1234" with 0.95 confidence). This dual-layered approach ensures systematic tie-breaking while emphasizing statistically robust reconstructions through confidence metrics and character prevalence.

These fusion strategies are adaptations of well-established methods in Pattern Recognition. For instance, MV-based fusion has demonstrated effectiveness in enhancing OCR performance for degraded documents [Reul *et al*., 2018], while HC is a widely adopted approach in speech recognition systems [Prabhavalkar *et al*., 2023].

4.3 Results

In this section, we present the results achieved by the OCR model in recognizing a single super-resolved LP image from each of the five SR baseline models. Following this, we analyze the employed fusion approaches and examine the impact of different resolutions on the results. Finally, we provide visual results for qualitative analysis.

In LPR research, model performance is traditionally evaluated using the recognition rate, defined as the ratio of correctly recognized LPs (where all characters are correctly classified) to the total number of LPs in the test set [Wang *et al*., 2022; Chen *et al*., 2023; Wei *et al*., 2024]. In addition to this metric, we incorporate partial matches (cases where at least 5 or 6 characters are recognized correctly). This approach is particularly valuable when not all LP characters can be accurately reconstructed or recognized, as it helps to narrow the search space.

Table 2 summarizes the performance of the GP_LPR model [Liu *et al*., 2024b] on both low- and high-resolution LP images, alongside images super-resolved by general-purpose methods SR3 [Saharia *et al*., 2023] and Real-ESRGAN [Wang *et al*., 2021], as well as LP-specialized networks LPSRGAN [Pan *et al*., 2024], PLNET [Nascimento *et al*., 2023, 2024a], and LCDNet [Nascimento *et al*., 2024b]. In LR samples, most characters are barely distinguishable, resulting in recognition rates as low as 2.2%.

The sharp decline in recognition rates for LR images underscores the challenges of achieving accurate LPR when working with low-quality inputs. While the GP_LPR model achieves 85.2% recognition rate on HR images, its performance plunges to just 2.2% on LR images, with only 20.1% of LPs having at least 5 correct characters. The ap-

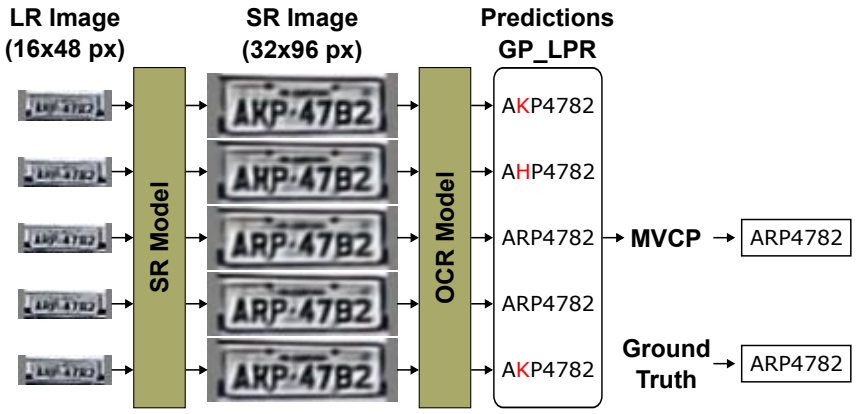

Figure 4. Illustration of the fusion process to enhance LPR performance by combining multiple super-resolved LP images. Sequential LR images (original size 16×48 pixels) from each track are independently upsampled to 32×96 pixels via a single-image SR model, then processed by the GP_LPR model [Liu *et al.*, 2024b]. The OCR outputs are aggregated using three fusion strategies: Highest Confidence (HC), Majority Vote (MV), and Majority Vote by Character Position (MVCP), leveraging temporal consistency across frames to resolve structural ambiguities in character reconstruction (e.g., distinguishing "R", "K" and "H"). This approach is particularly valuable for forensic and surveillance scenarios that demand high reliability.

Table 2. Recognition rates obtained by GP_LPR on different test images. The first LR and HR images from each track were used for the experiments on the original images. Super-resolved images were generated from the first LR image of each track.

Test Images	# Correct Characters		
	All	≥ 6	≥ 5
HR	85.2%	98.5%	99.8%
LR (no SR)	2.2%	8.2%	20.1%
LR + SR (SR3)	18.4%	45.8%	68.3%
LR + SR (LPSRGAN)	19.6%	46.1%	67.4%
LR + SR (Real-ESRGAN)	20.2%	49.8%	71.8%
LR + SR (PLNET)	29.9%	57.8%	76.9%
LR + SR (LCDNet)	29.9%	59.2%	77.1%

plication of SR methods, however, brings significant improvements. Domain-specific models such as LCDNet and PLNET achieve a 29.9% recognition rate, outperforming general-purpose SR approaches like SR3 (18.4%) and Real-ESRGAN (20.2%). For partial recognition (at least five correct characters), LCDNet reaches 77.1%, demonstrating that even imperfect reconstructions can effectively reduce the search space in forensic scenarios. Interestingly, LPSRGAN – despite being tailored for LPs – underperforms compared to Real-ESRGAN (19.6% vs. 20.2%), indicating potential limitations in its degradation modeling or training methodology. Overall, these findings highlight the promise of SR techniques in improving LPR robustness under real-world conditions. Nevertheless, the considerable gap between super-resolved (29.9%) and HR (85.2%) performance reveals the continued need for advancements in dealing with noise, compression, and other practical image degradations.

Table 3 presents the recognition rates achieved by combining the OCR model's predictions for three and five super-resolved images using the strategies described in Section 4.2. For LP-specialized models, PLNET achieves up to 40.9% recognition rate (all characters correct) with MVCP and 5 images, while LCDNet outperforms slightly at 42.3%, highlighting their robustness in reconstructing critical character details. Both models surpass general-purpose methods like SR3 (28.3%) and Real-ESRGAN (29.5%). The MVCP strategy consistently delivers the highest gains, improving PLNET's accuracy by 5.0% (3 to 5 images) and LCDNet's by 4.5% over HC and MV. Aggregating five images instead of three further boosts performance; for example, LCDNet achieves 86.9% for cases with at least 5 correct characters (vs. 83.3% with 3 images), nearing practical utility for surveillance systems. These results underscore the value of temporal fusion and domain-specific SR in overcoming real-world degradations, where partial matches (≥ 5 characters) remain critical for forensic tasks.

Tables 4 and 5 demonstrate the critical impact of image resolution on LP super-resolution and subsequent LPR performance. For lower-resolution sources (1280×960 pixels), PLNET and LCDNet achieve modest recognition rates of 22.7% and 23.5% (all characters correct, 5 images + MVCP), respectively. In contrast, for high-resolution sources (1920×1080 pixels), these models reach 59.3% (PLNET) and 61.4% (LCDNet) under the same conditions – a 2.6× improvement – underscoring the importance of pixel density in preserving structural details like character edges and serifs.

The MVCP fusion strategy consistently outperforms HC and MV across resolutions. For 1280×960 images, MVCP boosts LCDNet's accuracy by +5.0% (18.5% → 23.5%) with 5 images, while for 1920×1080 images, it improves results by +7.7% (53.1% → 61.4%). Increasing the number of fused images from 3 to 5 further enhances performance: partial matches (≥ 5 characters) rise from 78.2% to 95.8% for LCDNet in HR settings, demonstrating near-practical utility for surveillance systems.

Notably, even LR scenarios benefit significantly from fusion. For 1280×960 images, MVCP with 5 images achieves 71.6% (SR3) to 78.2% (LCDNet) for ≥ 5 correct characters, narrowing search spaces in forensic applications. This is critical for real-world deployments, where cost-effective cameras (e.g., 1280×960 sensors) dominate, and partial matches remain acceptable due to resolution constraints. The proposed dataset bridges this gap, enabling robust evaluation of SR methods across diverse operational scenarios.

While recent LP-specific SR methods (e.g., LPSRGAN [Pan *et al.*, 2024] and PLNET [Nascimento *et al.*, 2023, 2024a]) and general-purpose approaches (e.g., Real-ESRGAN [Wang *et al.*, 2021] and SR3 [Saharia *et al.*, 2023]) have advanced super-resolution research, our exper-

Table 3. Comparison of recognition rates achieved by combining the outputs of the GP_LPR model for multiple super-resolved images generated by PLNET and LCDNet. As outlined in Section 4.2, three fusion strategies were evaluated: Highest Confidence (HC), Majority Vote (MV), and Majority Vote by Character Position (MVCP).

Test Images (**Both Resolutions**)	# Images Majority Vote	HC			MV			MVCP		
		All	≥ 6	≥ 5	All	≥ 6	≥ 5	All	≥ 6	≥ 5
SR3 (LR + SR)	3	19.9%	49.5%	70.8%	21.4%	50.9%	71.2%	24.1%	56.3%	76.4%
	5	20.2%	50.6%	71.9%	25.1%	55.1%	73.8%	28.3%	61.8%	81.5%
LPSRGAN (LR + SR)	3	24.0%	51.5%	70.7%	24.5%	51.7%	70.9%	25.3%	53.5%	72.9%
	5	25.4%	53.1%	71.9%	27.4%	54.3%	73.0%	28.8%	56.3%	76.9%
Real-ESRGAN (LR + SR)	3	23.5%	54.6%	76.0%	24.1%	55.8%	76.2%	25.6%	57.4%	78.3%
	5	24.9%	56.6%	77.6%	27.7%	59.4%	78.5%	29.5%	61.7%	81.6%
PLNET (LR + SR)	3	34.5%	63.8%	80.7%	36.1%	64.4%	80.9%	36.6%	66.1%	82.0%
	5	35.9%	65.3%	82.9%	39.5%	67.4%	83.6%	40.9%	70.4%	85.8%
LCDNet (LR + SR)	3	34.8%	63.6%	81.9%	36.7%	64.5%	82.0%	37.8%	67.0%	83.3%
	5	35.7%	64.8%	83.1%	40.7%	68.1%	84.2%	42.3%	72.0%	86.9%

Table 4. Comparison of recognition rates achieved by combining the outputs of the GP_LPR model for multiple super-resolved images generated by PLNET and LCDNet (considering only LPs extracted from images with 1280×960 pixels). As outlined in Section 4.2, three fusion strategies were evaluated: Highest Confidence (HC), Majority Vote (MV), and Majority Vote by Character Position (MVCP).

Test Images (**1280×960**)	# Images Majority Vote	HC			MV			MVCP		
		All	≥ 6	≥ 5	All	≥ 6	≥ 5	All	≥ 6	≥ 5
SR3 (LR + SR)	3	9.5%	32.2%	56.8%	9.8%	33.3%	57.2%	11.8%	38.9%	63.9%
	5	9.9%	34.0%	58.4%	12.1%	37.1%	60.3%	15.3%	46.4%	71.6%
LPSRGAN (LR + SR)	3	8.7%	30.3%	52.2%	8.9%	30.3%	52.4%	8.7%	31.1%	55.9%
	5	9.6%	31.1%	54.0%	10.2%	32.3%	55.6%	11.0%	34.6%	60.9%
Real-ESRGAN (LR + SR)	3	12.3%	37.8%	63.7%	12.6%	39.0%	64.0%	13.7%	39.8%	65.8%
	5	14.0%	40.4%	65.6%	15.7%	42.5%	66.8%	17.4%	45.5%	71.1%
PLNET (LR + SR)	3	17.0%	45.5%	69.0%	17.7%	45.7%	69.3%	18.3%	47.3%	70.7%
	5	18.3%	47.0%	72.5%	20.7%	49.1%	73.7%	22.7%	53.4%	76.9%
LCDNet (LR + SR)	3	17.4%	44.3%	70.2%	18.5%	45.4%	70.5%	19.3%	48.4%	72.2%
	5	18.5%	46.2%	72.4%	21.7%	49.7%	73.9%	23.5%	55.3%	78.2%

Table 5. Comparison of recognition rates achieved by combining the outputs of the GP_LPR model for multiple super-resolved images generated by PLNET and LCDNet (considering only LPs extracted from images with 1920×1080 pixels). As outlined in Section 4.2, three fusion strategies were evaluated: Highest Confidence (HC), Majority Vote (MV), and Majority Vote by Character Position (MVCP).

Test Images (**1920×1080**)	# Images Majority Vote	HC			MV			MVCP		
		All	≥ 6	≥ 5	All	≥ 6	≥ 5	All	≥ 6	≥ 5
SR3 (LR + SR)	3	30.3%	66.8%	84.8%	33.0%	68.5%	85.3%	36.4%	73.8%	88.9%
	5	30.6%	67.3%	85.4%	38.2%	73.1%	87.4%	41.4%	77.2%	91.3%
LPSRGAN (LR + SR)	3	39.4%	72.7%	89.3%	40.2%	73.1%	89.4%	41.9%	76.0%	90.0%
	5	41.3%	75.1%	89.9%	44.7%	76.4%	90.5%	46.7%	78.1%	92.7%
Real-ESRGAN (LR + SR)	3	34.8%	71.5%	88.4%	35.7%	72.7%	88.5%	37.5%	75.1%	90.8%
	5	35.8%	72.8%	89.6%	39.8%	76.3%	90.3%	41.6%	78.0%	92.1%
PLNET (LR + SR)	3	52.2%	82.4%	92.5%	54.6%	83.3%	92.5%	55.1%	85.2%	93.5%
	5	53.7%	83.7%	93.2%	58.4%	85.9%	93.7%	59.3%	87.7%	94.9%
LCDNet (LR + SR)	3	52.3%	83.0%	93.6%	55.1%	83.8%	93.7%	56.4%	85.5%	94.5%
	5	53.1%	83.6%	93.8%	59.8%	86.7%	94.5%	61.4%	89.0%	95.8%

iments reveal their limitations in real-world scenarios. LP-SRGAN, despite its domain-specific design, achieves only 19.6% recognition accuracy on UFPR-SR-Plates (Table 2), a drastic drop from the 93.9% it attains on synthetic benchmarks like LicensePlateDataset10K [Pan *et al.*, 2024]. Similarly, PLNET, SR3, and LCDNet – which achieve 49.8%, 43.1%, and 39.0% on synthetic RodoSol-ALPR [Nascimento *et al.*, 2024b] – exhibit significantly lower accuracy on real-world UFPR-SR-Plates (29.9%, 18.4%, and 29.9%, respectively; Table 2). This disparity underscores the limitations of synthetic degradation pipelines, which fail to replicate real-world noise patterns (e.g., motion blur, sensor noise, weather effects). LCDNet's layout-aware architecture and OCR-guided training partially mitigate these challenges, achieving 61.4% accuracy with MVCP fusion (Table 3), but the gap between synthetic and real-world performance persists. These results emphasize the necessity of domain-specific architectures and real-world benchmarks like UFPR-SR-Plates to advance robust LPR systems.

Figures 5 and 6 present comparisons between low-resolution LP images and their super-resolved counterparts produced by the super-resolution networks. As expected, the GP_LPR model performs worse on low-resolution LPs (i.e., before super-resolution) extracted from 1280×960 im-

Figure 5. Recognition results for LPs cropped from images with a resolution of 1280 × 960 pixels. The top row shows the predictions made by GP_LPR [Liu *et al.*, 2024b] on the original LR images, while the two subsequent rows present the predictions obtained from super-resolved images generated by PLNET [Nascimento *et al.*, 2023, 2024a] and LCDNet [Nascimento *et al.*, 2024b]. Below each image, the predicted characters are displayed, with correct characters highlighted in blue and incorrect characters in red. The ground truth is indicated in green.

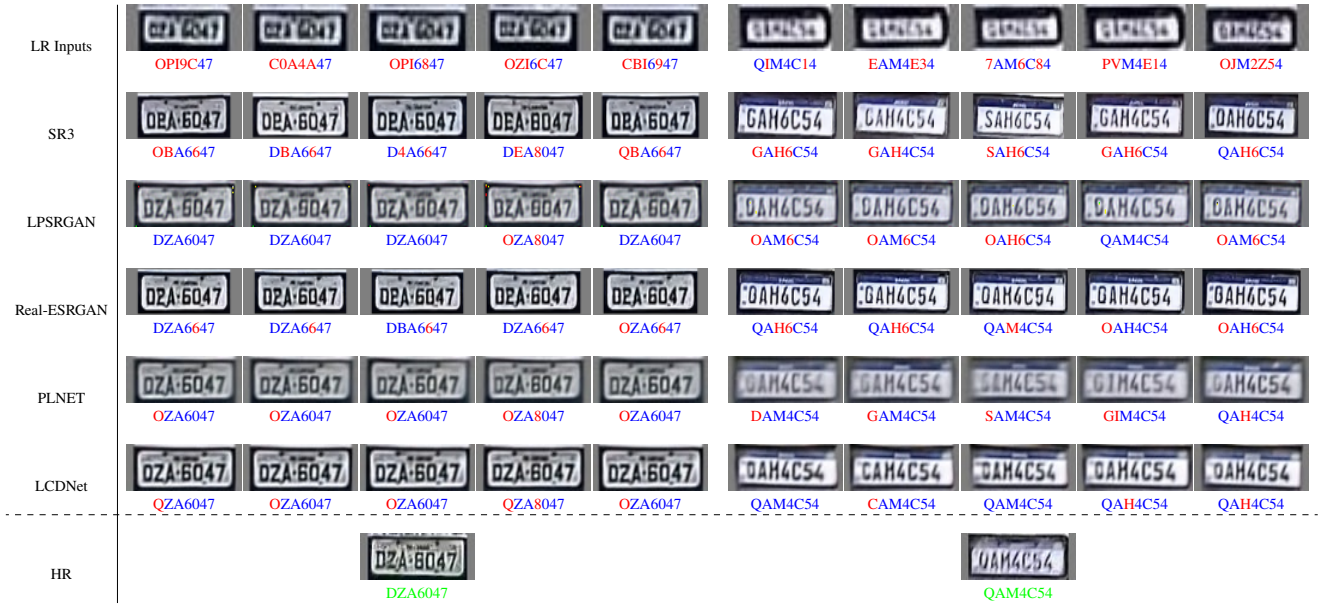

Figure 6. Recognition results for LPs cropped from images with a resolution of 1920 × 1080 pixels. The top row shows the predictions made by GP_LPR [Liu *et al.*, 2024b] on the original LR images, while the two subsequent rows present the predictions obtained from super-resolved images generated by PLNET [Nascimento *et al.*, 2023, 2024a] and LCDNet [Nascimento *et al.*, 2024b]. Below each image, the predicted characters are displayed, with correct characters highlighted in blue and incorrect characters in red. The ground truth is indicated in green.

ages (Figure 5) compared to 1920 × 1080 images (Figure 6). This performance gap stems primarily from the reduced pixel density in the smaller images, which causes characters to blend into the LP background. This effect is evident in the super-resolved LPs shown in Figure 5, where a "W" is mistakenly reconstructed as a "K" in the Brazilian LP (left), and a "B" is reconstructed as "R" or "H" in the Mercosur LP (right). Similar errors are observed in Figure 6, though less frequently due to the higher pixel density. For example, a "Q" was reconstructed as an "O," and an "M" was reconstructed as a "H," with PLNET-generated images exhibiting a more pronounced tendency towards these misclassifications than LCDNet-generated images.

4.3.1 Runtime Analysis

This section evaluates the inference efficiency and suitability of super-resolution models for both real-time and forensic applications. Table 6 presents a quantitative comparison of inference times and computational performance across the models. To ensure consistency and reliability, each model was independently tested over five separate runs.

LPSRGAN emerges as the fastest model, processing 207.5 frames per second (FPS) (4.82 ± 0.45 ms), making it well-suited for real-time applications such as traffic monitoring. However, this speed comes at the expense of performance: as shown in Table 3, the GP_LPR model achieves

Table 6. Inference Time and FPS Comparison.

Model	Time (ms) Avg ± Std	Time (ms) Min - Max	FPS Avg
LPSRGAN	4.82 ± 0.45	4.49 - 7.41	207.5
PLNET	24.88 ± 1.21	23.97 - 31.68	40.2
Real-ESRGAN	34.86 ± 2.66	32.53 - 46.46	28.7
LCDNet	61.88 ± 1.77	60.47 - 73.13	16.1

only a 28.8% recognition rate when applied to images super-resolved by LPSRGAN, considerably lower than the rates obtained with slower SR models. In contrast, LCDNet leads to the highest recognition rate (42.3% with MVCP fusion on 5 images), a 1.4× improvement over LPSRGAN, despite its slower inference speed (61.88 ms, 16.1 FPS). PLNET offers a balanced trade-off, combining moderate efficiency (24.88 ± 1.21 ms, 40.2 FPS) with competitive performance (40.9% recognition rate using MVCP), while Real-ESRGAN performs poorly in both inference speed (34.86 ms, 28.7 FPS) and recognition performance (29.5% recognition rate with MVCP).

Notably, LCDNet exhibits stable inference times (standard deviation: 1.77 ms), contrasting with LPSRGAN's higher variability (ranging from 4.49 to 7.41 ms). This stability, coupled with its superior accuracy under the MVCP protocol, establishes LCDNet as a strong candidate for forensic applications, where precision significantly outweighs speed requirements. These findings highlight a key trade-off: while latency-sensitive deployments may favor lightweight models such as LPSRGAN, accuracy-critical scenarios benefit more from specialized architectures like LCDNet, even at the cost of increased computational demands.

We also evaluated the SR3 model, which showed considerably higher inference times (12.31 ± 0.018 s). Although impractical for real-time use, SR3 may still be viable in forensic contexts where processing time is less restrictive. Due to its significantly slower performance – operating in seconds rather than milliseconds – SR3's results are omitted from Table 6, as they fall outside the real-time operational scope targeted by the other models.

5 Conclusions and Future Directions

In this work, we introduced UFPR-SR-Plates, a publicly available dataset specifically designed for LP super-resolution. The dataset comprises 100,000 LP images organized into 10,000 tracks, each containing five sequential LR and five sequential HR images of the same LP. This dataset is highly valuable for developing and evaluating super-resolution techniques aimed at improving License Plate Recognition (LPR) under real-world, low-quality conditions. Although primarily intended for LP super-resolution, UFPR-SR-Plates is also well-suited for training and evaluating LPR models, as it currently represents the largest collection of Brazilian and Mercosur LPs available.

We proposed a benchmark using the UFPR-SR-Plates dataset. We assessed the recognition rates achieved by the GP_LPR model [Liu *et al*., 2024b] in conjunction with five super-resolution approaches: general-purpose models (SR3 [Saharia *et al*., 2023] and Real-ESRGAN [Wang *et al*., 2021]) and LP-specialized networks (LPSRGAN, PLNET [Nascimento *et al*., 2023], and LCDNet [Nascimento *et al*., 2024b]). By fusing predictions from multiple super-resolved images via Majority Vote by Character Position (MVCP), recognition rates improved from 2.2% (raw LR images) to 29.9% for single-image outputs and up to 42.3% with five-image fusion using LCDNet. This represents a 19.2× improvement over raw LR images, with MVCP outperforming other fusion strategies by 3.88.1%, demonstrating its robustness in aggregating temporal information. Notably, LCDNet consistently surpassed both general-purpose and LP-specialized alternatives, validating the importance of architectural designs tailored to character structure and layout preservation.

Our findings offer valuable insights into the benefits of super-resolving multiple low-resolution versions of the same LP and combining their OCR predictions. This method is especially beneficial for surveillance and forensic applications, where partial matches can greatly reduce the search space for potential LPs. We believe this approach could be further strengthened by incorporating additional vehicle attributes, such as make, model, and color, as suggested by previous research [Oliveira *et al*., 2021; Lima *et al*., 2024].

For future work, we aim to enhance LP domain-specific architectures by incorporating multi-image fusion to enable temporal learning. Additionally, we plan to expand the UFPR-SR-Plates dataset to address its current limitations: (i) Motorcycle LPs, which were excluded due to their two-row layout and acquisition challenges – such as the relatively low traffic of motorcycles at our capture site and frequent repetitions of the same LPs; (ii) Systematic OCR errors (e.g., confusion between "B" and "8"), which we intend to mitigate using multi-model ensembling or confidence calibration techniques; (iii) Regional and nighttime LPs, as the current dataset is limited to daylight Brazilian/Mercosur LPs – we aim to align it with recent multi-region Automatic License Plate Recognition (ALPR) systems [Laroca *et al*., 2022]; (iv) Generalization to extreme conditions (e.g., rain, haze) via hybrid synthetic-real training.

We remark that these limitations do not undermine UFPR-SR-Plates's value as a foundational benchmark for real-world SR research. On the contrary, these limitations underscore clear opportunities for advancing robust ALPR systems. By publicly releasing UFPR-SR-Plates, we aim to drive progress in super-resolution and recognition of degraded LPs, particularly in unconstrained scenarios where existing synthetic benchmarks prove insufficient.

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 gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro RTX 8000 GPU used for this research.

Declarations

Authors' Contributions

Valfride Nascimento is the main contributor and writer of this manuscript. Gabriel E. Lima contributed to experiment validation and manuscript review. Rafael O. Ribeiro provided project resources and contributed to the manuscript review process. William R. Schwartz co-supervised the project and reviewed the manuscript. Rayson Laroca contributed to the conceptualization and methodology and assisted in the review and editing of the manuscript. David Menotti supervised the project, provided project resources, and contributed to the review and editing of the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The resources generated and analyzed during this study are available at `https://valfride.github.io/nascimento2025toward/`.

References

AlHalawani, S., Benjdira, B., Ammar, A., Koubaa, A., and Ali, A. M. (2024). DiffPlate: A diffusion model for super-resolution of license plate images. *Electronics*, 13(13):2670. DOI: 10.3390/electronics13132670.

Angelika Mulia, D., Safitri, S., and Gede Putra Kusuma Negara, I. (2024). YOLOv8 and Faster R-CNN performance evaluation with super-resolution in license plate recognition. *International Journal of Computing and Digital Systems*, 16(1):365–375. DOI: 10.12785/ijcds/160129.

Chen, S.-L., Liu, Q., Chen, F., and Yin, X.-C. (2023). End-to-end multi-line license plate recognition with cascaded perception. In *International Conference on Document Analysis and Recognition (ICDAR)*, pages 274–289. DOI: 10.1007/978-3-031-41734-4_17.

Ghamrawi, N. and McCallum, A. (2005). Collective multi-label classification. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 195–200. DOI: 10.1145/1099554.109959.

Gonçalves, G. R. *et al*. (2018). Real-time automatic license plate recognition through deep multi-task networks. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 110–117. DOI: 10.1109/SIBGRAPI.2018.00021.

Gonçalves, G. R. *et al*. (2019). Multi-task learning for low-resolution license plate recognition. In *Iberoamerican Congress on Pattern Recognition (CIARP)*, pages 251–261. DOI: 10.1007/978-3-030-33904-3_23.

Hamdi, A., Chan, Y. K., and Koo, V. C. (2021). A new image enhancement and super resolution technique for license plate recognition. *Heliyon*, 7(11). DOI: 10.1016/j.heliyon.2021.e08341.

Hijji, M., Khan, A., Alwakeel, M. M., Harrabi, R., Aradah, F., Cheikh, F. A., Sajjad, M., and Muhammad, K. (2023). Intelligent image super-resolution for vehicle license plate in surveillance applications. *Mathematics*, 11(4):892. DOI: 10.3390/math11040892.

Jia, W. and Xie, M. (2023). An efficient license plate detection approach with deep convolutional neural networks in unconstrained scenarios. *IEEE Access*, 11:85626–85639. DOI: 10.1109/ACCESS.2023.3301122.

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV)*, pages 694–711. DOI: 10.1007/978-3-319-46475-6_43.

Ke, X. *et al*. (2023). An ultra-fast automatic license plate recognition approach for unconstrained scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 24(5):5172–5185. DOI: 10.1109/TITS.2023.3237581.

Kim, D., Kim, J., and Park, E. (2024). AFA-Net: Adaptive feature attention network in image deblurring and super-resolution for improving license plate recognition. *Computer Vision and Image Understanding*, 238:103879. DOI: 10.1016/j.cviu.2023.103879.

Kittler, J., Hatef, M., Duin, R. P., and Matas, J. (1998). On combining classifiers. *IEEE transactions on pattern analysis and machine intelligence*, 20(3):226–239. DOI: 10.1109/34.667881.

Laroca, R., dos Santos, M., and Menotti, D. (2025). Improving small drone detection through multi-scale processing and data augmentation. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8.

Laroca, R., Estevam, V., Britto Jr., A. S., Minetto, R., and Menotti, D. (2023a). Do we train on test data? The impact of near-duplicates on license plate recognition. In *International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. DOI: 10.1109/IJCNN54540.2023.10191584.

Laroca, R., Zanlorensi, L. A., Estevam, V., Minetto, R., and Menotti, D. (2023b). Leveraging model fusion for improved license plate recognition. In *Iberoamerican Congress on Pattern Recognition*, pages 60–75. DOI: 10.1007/978-3-031-49249-5_5.

Laroca, R., Zanlorensi, L. A., Gonçalves, G. R., Todt, E., Schwartz, W. R., and Menotti, D. (2021). An efficient and layout-independent automatic license plate recognition system based on the YOLO detector. *IET Intelligent Transport Systems*, 15(4):483–503. DOI: 10.1049/itr2.12030.

Laroca, R. *et al*. (2022). On the cross-dataset generalization in license plate recognition. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 166–178. DOI: 10.5220/0010846800003124.

Ledig, C. *et al*. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114. DOI: 10.1109/CVPR.2017.19.

Lee, S., Kim, J.-H., and Heo, J.-P. (2020). Super-resolution of license plate images via character-based perceptual loss. In *IEEE International Conference on Big Data and Smart Computing (BigComp)*, pages 560–563. DOI: 10.1109/BigComp48618.2020.000-1.

Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., and Timofte, R. (2021). Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1833–1844. DOI: 10.1109/ICCVW54120.2021.00210.

Lima, G. E. *et al*. (2024). Toward enhancing vehicle color recognition in adverse conditions: A dataset and benchmark. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 1–6. DOI: 10.1109/SIBGRAPI62404.2024.10716307.

Lin, M., Liu, L., Wang, F., Li, J., and Pan, J. (2021). License plate image reconstruction based on generative adversarial networks. *Remote Sensing*, 13(15):3018. DOI: 10.3390/rs13153018.

Liu, A., Liu, Y., Gu, J., Qiao, Y., and Dong, C. (2023). Blind image super-resolution: A survey and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5461–5480. DOI: 10.1109/TPAMI.2022.3203009.

Liu, Q., Liu, Y., Chen, S.-L., Zhang, T.-H., Chen, F., and Yin, X.-C. (2024a). Improving multi-type license plate recognition via learning globally and contrastively. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–11. Early Access. DOI: 10.1109/TITS.2024.3365537.

Liu, Y.-Y., Liu, Q., Chen, S.-L., Chen, F., and Yin, X.-C. (2024b). Irregular license plate recognition via global information integration. In *International Conference on Multimedia Modeling*, pages 325–339. DOI: 10.1007/978-3-031-53308-2_24.

Luo, X., Huang, Y., and Miao, W. (2024). Real-world license plate image super-resolution via domain-specific degradation modeling. In *IEEE Conference on Artificial Intelligence (CAI)*, pages 1175–1180. IEEE. DOI: 10.1109/CAI59869.2024.00210.

Maier, A., Moussa, D., Spruck, A., Seiler, J., and Riess, C. (2022). Reliability scoring for the recognition of degraded license plates. In *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 1–8. DOI: 10.1109/AVSS56176.2022.9959390.

Mehri, A., Ardakani, P. B., and Sappa, A. D. (2021). MPRNet: Multi-path residual network for lightweight image super resolution. In *IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 2703–2712. DOI: 10.1109/WACV48630.2021.00275.

Moussa, D. *et al*. (2022). Forensic license plate recognition with compression-informed transformers. In *IEEE International Conference on Image Processing (ICIP)*, pages 406–410. DOI: 10.1109/ICIP46576.2022.9897178.

Nascimento, V., Laroca, R., and Menotti, D. (2024a). Super-resolution towards license plate recognition. In *Anais do XXXVII Concurso de Teses e Dissertações (CTD) do Congresso da Sociedade Brasileira de Computação (CSBC)*, pages 78–87. DOI: 10.5753/ctd.2024.1999.

Nascimento, V., Laroca, R., Ribeiro, R. O., Schwartz, W. R., and Menotti, D. (2024b). Enhancing license plate super-resolution: A layout-aware and character-driven approach. *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 1–6. DOI: 10.1109/SIBGRAPI62404.2024.10716303.

Nascimento, V. *et al*. (2022). Combining attention module and pixel shuffle for license plate super-resolution. In *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 228–233. DOI: 10.1109/SIBGRAPI55357.2022.9991753.

Nascimento, V. *et al*. (2023). Super-resolution of license plate images using attention modules and sub-pixel convolution layers. *Computers & Graphics*, 113:69–76. DOI: 10.1016/j.cag.2023.05.005.

Oliveira, I. O. *et al*. (2021). Vehicle-Rear: A new dataset to explore feature fusion for vehicle identification using convolutional neural networks. *IEEE Access*, 9:101065–101077. DOI: 10.1109/ACCESS.2021.3097964.

Pan, S., Chen, S.-B., and Luo, B. (2023). A super-resolution-based license plate recognition method for remote surveillance. *Journal of Visual Communication and Image Representation*, 94:103844. DOI: 10.1016/j.jvcir.2023.103844.

Pan, Y., Tang, J., and Tjahjadi, T. (2024). LPSRGAN: Generative adversarial networks for super-resolution of license plate image. *Neurocomputing*, 580:127426. DOI: 10.1016/j.neucom.2024.127426.

Prabhavalkar, R., Hori, T., Sainath, T. N., Schlüter, R., and Watanabe, S. (2023). End-to-end speech recognition: A survey. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32:325–351. DOI: 10.1109/TASLP.2023.3328283.

Presidência da República (2014). Lei nº 9.503, de 23 de setembro de 1997. Código de Trânsito Brasileiro. http://www.planalto.gov.br/ccivil_03/leis/l9503compilado.htm.

Rao, Z. *et al*. (2024). License plate recognition system in unconstrained scenes via a new image correction scheme and improved CRNN. *Expert Systems with Applications*, 243:122878. DOI: 10.1016/j.eswa.2023.122878.

Redmon, J., Divvala, S., Girshick, R., and Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788. DOI: 10.1109/CVPR.2016.91.

Reul, C., Springmann, U., Wick, C., and Puppe, F. (2018). Improving ocr accuracy on early printed books by utilizing cross fold training and voting. In *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, pages 423–428. IEEE. DOI: 10.1109/DAS.2018.30.

Saharia, C., Ho, J., Chan, W., Salimans, T., Fleet, D. J., and Norouzi, M. (2023). Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726. DOI: 10.1109/TPAMI.2022.3204461.

Schirrmacher, F., Lorch, B., Maier, A., and Riess, C. (2023). Benchmarking probabilistic deep learning methods for license plate recognition. *IEEE Transactions on Intelligent Transportation Systems*, 24(9):9203–9216. DOI: 10.1109/TITS.2023.3278533.

Sendjasni, A. and Larabi, M.-C. (2024). Embedding similarity learning for extreme license plate super-resolution. In *2024 IEEE 26th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE. DOI: 10.1109/MMSP61759.2024.10743401.

Shi, B., Yang, M., Wang, X., Lyu, P., Yao, C., and Bai, X. (2019). ASTER: An attentional scene text recognizer with flexible rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9):2035–2048. DOI: 10.1109/TPAMI.2018.2848939.

Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Rueckert, D., and Wang, Z. (2016). Real-time single image and video super-resolution using an

efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883. DOI: 10.48550/arXiv.1609.05158.

Silva, S. M. and Jung, C. R. (2020). Real-time license plate detection and recognition using deep convolutional neural networks. *Journal of Visual Communication and Image Representation*, page 102773. DOI: 10.1016/j.jvcir.2020.102773.

Silva, S. M. and Jung, C. R. (2022). A flexible approach for automatic license plate recognition in unconstrained scenarios. *IEEE Transactions on Intelligent Transportation Systems*, 23(6):5693–5703. DOI: 10.1109/TITS.2021.3055946.

Ultralytics (2023). YOLOv8. Accessed: 2024-06-28.

Wang, X., Xie, L., Dong, C., and 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*, pages 1905–1914. DOI: 10.1109/ICCVW54120.2021.00217.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0. DOI: 10.1007/978-3-030-11021-5_5.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Loy, C. C. (2019). ESRGAN: Enhanced super-resolution generative adversarial networks. In *European Conference on Computer Vision (ECCV)*, pages 63–79. DOI: 10.1007/978-3-030-11021-5_5.

Wang, Y., Bian, Z.-P., Zhou, Y., and Chau, L.-P. (2022). Rethinking and designing a high-performing automatic license plate recognition approach. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):8868–8880. DOI: 10.1109/TITS.2021.3087158.

Wei, C., Han, F., Fan, Z., Shi, L., and Peng, C. (2024). Efficient license plate recognition in unconstrained scenarios. *Journal of Visual Communication and Image Representation*, 104:104314. DOI: 10.1016/j.jvcir.2024.104314.

Xu, Z., Yang, W., Meng, A., Lu, N., Huang, H., Ying, C., and Huang, L. (2018). Towards end-to-end license plate detection and recognition: A large dataset and baseline. In *European Conference on Computer Vision (ECCV)*, pages 261–277. DOI: 10.1007/978-3-030-01261-8_16.

Zhang, R. *et al*. (2018). The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 586–595. DOI: 10.1109/CVPR.2018.00068.

Zhao, H.-h. and Liu, H. (2020). Multiple classifiers fusion and cnn feature extraction for handwritten digits recognition. *Granular Computing*, 5(3):411–418. DOI: 10.1007/s41066-019-00158-6.

Zhou, Z.-H. (2025). *Ensemble methods: foundations and algorithms*. CRC press.