























# ICPR 2026 Competition on Low-Resolution License Plate Recognition

Rayson Laroca<sup>1,2,\*</sup> , Valfride Nascimento<sup>1</sup> , Donggun Kim<sup>3</sup> ,  
Sanghyeok Chung<sup>3</sup> , Subin Bae<sup>3</sup> , Uihwan Seo<sup>3</sup> , Seungsang Oh<sup>3</sup> ,  
Chi M. Phung<sup>4,6</sup> , Minh G. Vo<sup>5,6</sup> , Xingsong Ye<sup>7,8</sup> , Yongkun Du<sup>7,8</sup> ,  
Yuchen Su<sup>7,8</sup> , Zhineng Chen<sup>7,8</sup> , Sunhee Heo<sup>9</sup> , Hyangwoo Lee<sup>9</sup> ,  
Kihyun Na<sup>9</sup> , Khanh V. Vu Nguyen<sup>4,6</sup> , Sang T. Pham<sup>4,6</sup> ,  
Duc N. N. Phung<sup>4,6</sup> , Trong P. Le<sup>4,6</sup> , Vy N. Vo Tran<sup>4,6</sup> , and  
David Menotti<sup>1</sup> 

<sup>1</sup> Federal University of Paraná, Brazil

<sup>2</sup> Pontifical Catholic University of Paraná, Brazil

<sup>3</sup> Korea University, Republic of Korea

<sup>4</sup> University of Information Technology, Vietnam

<sup>5</sup> Ho Chi Minh University of Technology, Vietnam

<sup>6</sup> Vietnam National University, Vietnam

<sup>7</sup> Fudan University, China

<sup>8</sup> Shanghai Key Laboratory of Multimodal Embodied AI, China

<sup>9</sup> Handong Global University, Republic of Korea

\* rayson@ppgia.pucpr.br

**Abstract.** Low-Resolution License Plate Recognition (LRLPR) remains a challenging problem in real-world surveillance scenarios, where long capture distances, compression artifacts, and adverse imaging conditions can severely degrade license plate legibility. To promote progress in this area, we organized the ICPR 2026 Competition on Low-Resolution License Plate Recognition, the first competition specifically dedicated to LRLPR using real low-quality data collected under operationally relevant conditions. The competition was based on the LRLPR-26 dataset, which comprises 20,000 training tracks and 3,000 test tracks; each training track contains five low-resolution and five high-resolution images of the same license plate. Notably, a total of 269 teams from 41 countries registered for the competition, and 99 teams submitted valid entries in the Blind Test Phase. The winning team achieved a Recognition Rate of 82.13%, and four teams surpassed the 80% mark, highlighting both the high level of competition at the top of the leaderboard and the continued difficulty of the task. In addition to presenting the competition design, evaluation protocol, and main results, this paper summarizes the methods adopted by the top-5 teams and discusses current trends and promising directions for future research on LRLPR. The competition webpage is available at <https://icpr26lrlpr.github.io/>

**Keywords:** Automatic license plate recognition · Low-resolution · Real-world scenarios · Forensics · Challenge.

## 1 Introduction

Automatic License Plate Recognition (ALPR) systems rely on image processing and pattern recognition techniques to detect and recognize License Plates (LPs) in images or videos. They are widely used in traffic law enforcement, electronic toll collection, and vehicle access control in restricted areas [23, 63].

With the evolution of general-purpose object detectors, particularly YOLO and its variants [59], LP detection performance has approached saturation under standard imaging conditions [26, 33, 56]. Recent studies have also reported very high recognition rates for complete ALPR pipelines [32, 34, 51]. However, such results are largely obtained from high-quality images, where LP characters are sharp, well-defined, and minimally affected by noise or compression artifacts.

In real-world surveillance environments, LP images are frequently captured at low resolution due to hardware constraints or the large distance between vehicles and cameras [15, 43, 50]. In addition, storage and bandwidth limitations often require strong compression, further degrading visual quality [45, 53, 64]. As a result, characters may appear blurred, distorted, or barely distinguishable from the background, significantly increasing recognition difficulty [13, 14, 42].

Despite its clear practical importance, Low-Resolution License Plate Recognition (LRLPR) remains a challenging and relatively underexplored problem with strong forensic and societal relevance [41, 53]. Even state-of-the-art approaches struggle to exceed 50–60% recognition accuracy on real-world low-quality images [15, 44, 45]. Enhancing recognition performance under such adverse conditions can significantly reduce investigative search spaces and support faster, more reliable decision-making in law enforcement workflows [39, 53].

Although several studies report promising results on LRLPR [27, 41, 43], most rely on synthetically degraded images derived from high-resolution samples, typically using bicubic downsampling. Such simplified degradation models fail to capture the complex artifacts and variability present in real operational scenarios [14, 15, 42]. This limitation underscores the importance of benchmarks built from genuinely low-quality data collected under real-world conditions.

To foster progress in this direction, we organized the first *Competition on Low-Resolution License Plate Recognition*, held in conjunction with the 2026 International Conference on Pattern Recognition (ICPR)<sup>1</sup>. The competition was based on the LRLPR-26 dataset, an expanded version of our previously released benchmark [45]. The dataset comprises 200,000 images organized into 20,000 training tracks, each containing five low-resolution (LR) images and five high-resolution (HR) images of the same LP, as well as 15,000 images organized into 3,000 test tracks, each containing five LR images of the same LP. To the best of our knowledge, LRLPR-26 is the largest public dataset featuring real low- and high-resolution LPs acquired from the same vehicles.

Remarkably, the competition attracted 269 teams from 41 countries worldwide, which submitted 99 entries during the final evaluation phase alone. In this paper, we present an overview of the competition, including the dataset design,

<sup>1</sup> <https://icpr26lrlpr.github.io/>

the evaluation protocol, the results, and a description of the five top-ranked approaches. Our goal is to foster further advances in LRLPR and to contribute to the broader development of scene text recognition systems that can operate reliably under adverse imaging conditions.

The rest of this paper is organized as follows. Section 2 presents the competition, including the dataset, competition phases, evaluation protocol, participation statistics, and fairness rules. Section 3 reports the competition results and summarizes the top-ranked approaches. Section 4 discusses the main findings and implications of the competition. Finally, Section 5 concludes the paper.

## 2 The 2026 LRLPR Competition

The competition was designed with two main objectives. First, it aimed to provide a broad view of recent advances and emerging trends in LRLPR. Second, it sought to bring together researchers from different institutions and backgrounds, fostering exchange and potential collaboration within the research community.

The remainder of this section is organized as follows. Sections 2.1 and 2.2 describe the training and test data provided to participants. Section 2.3 discusses privacy considerations associated with the dataset. Section 2.4 presents the competition phases and submission format, and Section 2.5 details the evaluation protocol. Section 2.6 summarizes participation statistics, while Section 2.7 outlines the main rules adopted to ensure a fair and transparent competition.

### 2.1 Training Data

The training data comprises 20,000 tracks, each containing five consecutive low-resolution (LR) images and five consecutive high-resolution (HR) images of the same LP, for a total of 200,000 images. Fig. 1 shows all LP images from four example tracks. The inclusion of HR images in the training set was intended to encourage participants to explore image enhancement strategies, such as super-resolution, that could improve recognition performance.



**Fig. 1.** Examples of tracks from the training data. Each row corresponds to a complete track, showing five consecutive low-resolution (LR) images on the left and five consecutive high-resolution (HR) captures of the same LP on the right.

The original videos (before LP cropping) were acquired with a rolling-shutter camera installed at the Federal University of Paraná, in Curitiba, Brazil, under

conditions designed to resemble real-world surveillance scenarios. Half of the training set, corresponding to 10,000 tracks, comes from the recently published UFPR-SR-Plates dataset [45]. These tracks, hereinafter referred to as Scenario A, were collected under relatively controlled conditions, such as daylight and no rain. Although additional details about this dataset are available in [45], its acquisition procedure was essentially the same as that adopted for the remaining 10,000 tracks, described next as Scenario B.

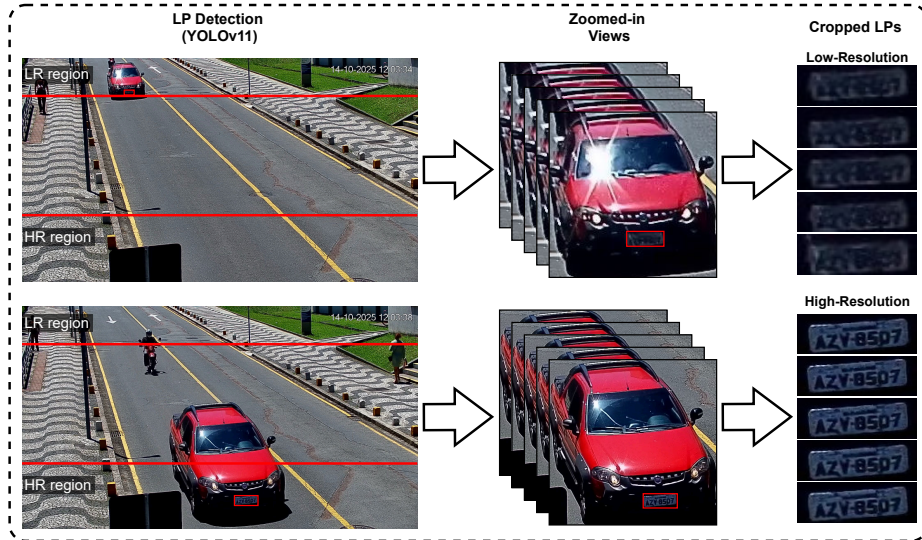
Scenario B was collected specifically for this competition. The same camera used in Scenario A was employed, but it was oriented in a different direction, and the resulting data cover a broader range of environmental conditions, including rain and nighttime. Fig. 2 shows representative images used to build the dataset (i.e., prior to LP detection and extraction), highlighting the diversity of vehicle categories and the variety of environmental conditions.



**Fig. 2.** Representative full-frame images (i.e., before LP detection and cropping) used to construct the LRLPR-26 dataset. The top row corresponds to Scenario A, whereas the bottom row corresponds to Scenario B, illustrating variations in vehicle categories and environmental conditions, including daylight, rain, and nighttime.

The cropped LPs were obtained from video sequences of vehicles entering and leaving the roadway in opposite directions, as illustrated in Fig. 3. All videos were recorded at a resolution of  $1920 \times 1080$  pixels. To detect and track LPs, we employed YOLOv11 [60], motivated by the strong performance of the YOLO family in unconstrained scenarios [8, 29, 35]. The detector was fine-tuned on widely used datasets in the ALPR literature, including RodoSol-ALPR [28] and UFPR-ALPR [31]. Detected LPs were then associated across frames using BoT-SORT [1]. From each track, the five patches extracted from the frames farthest from the camera were selected as the LR samples, whereas the five patches extracted from the nearest frames were selected as the HR samples. To annotate the LP text, we applied the multi-task Optical Character Recognition (OCR) model proposed by Gonçalves et al. [12] to each of the five HR images in a track and obtained the final label through majority voting [32].

The dataset contains two LP layouts, namely Brazilian and Mercosur. Brazilian LPs follow the pattern of three letters followed by four digits, whereas Mercosur LPs in Brazil follow the pattern of three letters, one digit, one letter, and two digits [28]. The distribution of tracks for each layout across Scenarios A and B



**Fig. 3.** Overview of the dataset construction pipeline. LPs are first detected using YOLOv11 [60] and then tracked across frames with BoT-SORT [1]. For each vehicle, patches extracted from frames farther from the camera are used as low-resolution (LR) samples, whereas patches from frames closer to the camera are used as high-resolution (HR) samples. The final LP transcription is then obtained semi-automatically by applying an OCR model [12] to the HR patches.

is detailed in Table 1. The higher concentration of Mercosur LPs in Scenario B reflects their current prevalence in the Brazilian automotive fleet. Furthermore, Scenario B was designed with a strictly unique mapping—one track per LP (and thus per vehicle)—thereby increasing dataset diversity compared to Scenario A.

**Table 1.** Distribution of training tracks by LP layout.

	Number of Tracks			Unique LPs
	Brazilian	Mercosur	Total	
Scenario A [45]	5,000	5,000	10,000	6,946
Scenario B (new)	2,000	8,000	10,000	10,000

Each track in Scenario A is annotated with the LP text, the LP layout, and the  $(x, y)$  coordinates of the four LP corners [45]. In Scenario B, however, corner annotations were not provided due to the high cost of manual labeling. Although much of the acquisition pipeline was automated, all annotations were manually reviewed to ensure the reliability of the competition data. During this process, occasional errors, mainly introduced by the OCR model, were identified and corrected.

Participants were free to determine how much of the training data to reserve for validation. As discussed in Section 3.1, the top-ranked teams adopted different validation strategies. They were also free to choose how to combine predictions from the five LR images, for example, through majority voting, confidence-based selection, or temporal modeling.

## 2.2 Test Data

The test set comprises 3,000 tracks, all collected exclusively from Scenario B, with each track corresponding to a unique vehicle. None of the LPs in the test set appears in the training data. Its layout distribution matches that of the training set, comprising 600 tracks with Brazilian LPs and 2,400 with Mercosur LPs.

The test set does not include annotations or HR images, as participants were required to predict the LP text using only the provided LR images.

## 2.3 Privacy Concerns

We remark that LPs of vehicles registered in Brazil are not publicly linked to the personal information of their owners, which substantially reduces privacy risks. An LP uniquely identifies the vehicle itself and does not, by design, disclose personal data. Accordingly, the LP remains associated with the vehicle even after a transfer of ownership, underscoring its function as a vehicle identifier rather than a personal identifier. This characteristic has been previously discussed in prior studies introducing datasets collected in Brazil [28, 35, 45, 46].

## 2.4 Competition Phases and Submission Format

Although the competition had a single track, it was conducted in two phases.

The first phase, referred to as the *Public Test Phase*, lasted approximately one month. During this stage, participants had access to a subset of 1,000 out of the 3,000 test tracks, of which 200 corresponded to Brazilian LPs and 800 to Mercosur LPs. This phase allowed participants to obtain feedback on their methods through Codabench [66] using part of the test set, while a public leaderboard was maintained on the same platform. Each team was allowed up to five submissions per day, with a maximum of 25 submissions overall.

Immediately after the end of the *Public Test Phase*, the competition moved to its final stage, called the *Blind Test Phase*, which lasted approximately one week. At that point, the full test set, comprising the 3,000 tracks described in Section 2.2, was made available. During this phase, the leaderboard became private: each team could view only its own score, but not the scores of the other teams. To reduce leaderboard probing, each team was allowed up to three submissions in total, which also provided some margin in case of platform-related issues.

In both phases, participants were required to submit a `.txt` prediction file in which each line consisted of `track_id,plate_text;confidence`, as illustrated below:

```

track_00001,ABC1234;0.9876
track_00002,DEF5678;0.6789
track_00003,GHI9012;0.4521
...

```

The predicted LP strings and their corresponding confidence scores were used to calculate the evaluation metrics detailed in the following section.

## 2.5 Evaluation Protocol

The official ranking was determined by the *Recognition Rate* computed on the Blind Test Set. This metric evaluates performance at the track level and requires an exact match between the predicted and ground-truth LP strings [30, 56, 62].

Formally, the Recognition Rate is defined as:

$$\text{Recognition Rate} = \frac{\text{Number of Correctly Recognized Tracks}}{\text{Number of Tracks in the Test Set}}. \quad (1)$$

When two or more teams achieved the same Recognition Rate, ties were broken using a secondary criterion, *Confidence Gap*, which measures the model’s ability to distinguish its correct from incorrect predictions based on its confidence scores. It is defined as the difference between the mean confidence assigned to correct predictions and that assigned to incorrect predictions. A higher Confidence Gap indicates better-calibrated confidence estimates.

The results reported in this paper are primarily based on the submissions made to the Codabench server<sup>2</sup>. A script for computing these metrics is available at <https://raysonlaroca.github.io/supp/lr1pr26/>.

## 2.6 Participants

The competition attracted 269 teams from 41 countries, with the largest contingents coming from Vietnam, India, Mainland China, Taiwan, and the Republic of Korea. Among these, 118 teams submitted valid entries in the Public Test Phase, while 99 teams participated in the Blind Test Phase. This level of engagement, specifically acknowledged in official communications from the Codabench team, is particularly remarkable for a competition without monetary prizes.

The competition would certainly have attracted even more teams if registration had been open to all interested participants. However, registration and, consequently, access to the dataset were restricted to teams affiliated with accredited universities or research institutions. This follows a common access policy adopted by several datasets in the ALPR literature [20, 28, 31], including UF-PR-SR-Plates [45], which was incorporated into the official training data.

<sup>2</sup> <https://www.codabench.org/competitions/12259/>

## 2.7 Fairness

To ensure a fair and transparent competition, a set of rules was established, following principles similar to those adopted in other recent competitions [6, 7]. Among the most important were the following. Manual inspection or annotation of the test set was strictly prohibited. At the same time, participants were allowed to train their methods using additional datasets, such as UFPR-ALPR [31] and RodoSol-ALPR [28]. These rules were designed to discourage test-set leakage while still allowing methodological flexibility.

## 3 Results

The results of the top 20 teams are presented in Table 2, while Figs. 4 and 5 provide a broader view of the performance distribution across all participating teams. The complete leaderboard is available on the competition webpage.

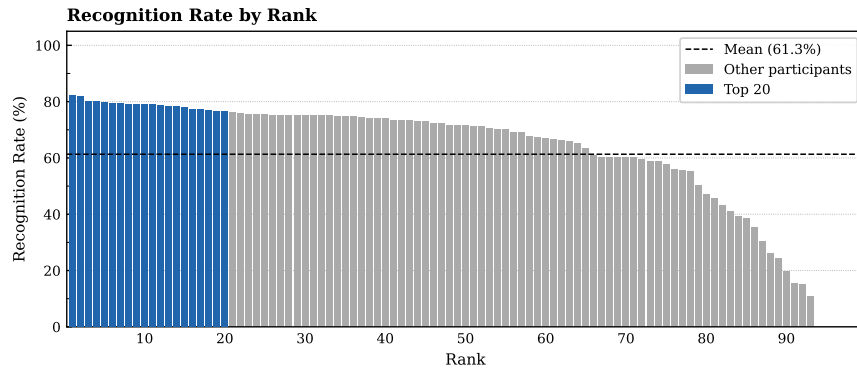
**Table 2.** Top 20 teams in the competition, ranked according to the official criterion, namely the Recognition Rate computed on the Blind Test Set. Confidence Gap is also reported, as it was used as a secondary criterion to break ties when needed.

Rank	Recognition Rate ( $\uparrow$ )	Confidence Gap ( $\uparrow$ )	Rank	Recognition Rate ( $\uparrow$ )	Confidence Gap ( $\uparrow$ )
1	82.13%	6.67%	11	79.00%	3.67%
2	81.73%	3.75%	12	78.60%	8.58%
3	80.17%	2.38%	13	78.23%	12.15%
4	80.10%	14.86%	14	78.20%	4.42%
5	79.83%	5.93%	15	77.97%	11.38%
6	79.50%	20.47%	16	77.37%	6.94%
7	79.23%	12.37%	17	77.30%	22.80%
8	79.13%	6.77%	18	77.07%	12.63%
9	79.10%	6.55%	19	76.57%	8.04%
10	79.00%	13.92%	20	76.47%	13.04%

Table 2 shows that the competition was highly competitive at the top of the leaderboard. The winning team achieved a Recognition Rate of 82.13%, followed by 81.73% and 80.17% for the 2<sup>nd</sup>- and 3<sup>rd</sup>-ranked teams, respectively. Notably, four teams surpassed the 80% mark, and the top 10 teams were separated by only 3.13 percentage points. Even when considering the entire top 20, the gap between the 1<sup>st</sup>- and 20<sup>th</sup>-ranked teams was only 5.66 percentage points, from 82.13% to 76.47%. This narrow spread indicates that relatively small performance gains were sufficient to produce substantial changes in the final ranking.

This strong competition at the top becomes even clearer in Fig. 4, which places the leading teams in the context of the full leaderboard. All top 20 teams achieved a Recognition Rate of more than 76%, whereas the overall mean across

participants was 61.3%. The leftmost portion of the curve is notably flat, indicating that the highest-ranked teams were separated by relatively small margins despite achieving very strong performance. More specifically, the recognition rate declines only slowly from the top positions through much of the ranking, and a substantial number of teams remained above the overall mean. This pattern suggests that many teams were able to develop reasonably effective solutions, even if only a smaller group reached the highest performance band. At the same time, Fig. 4 underscores that LRLPR is far from a solved problem. Under the strict exact-match evaluation protocol standard in ALPR research [30, 56, 62], even the top-performing method failed on 17.87% of test tracks.



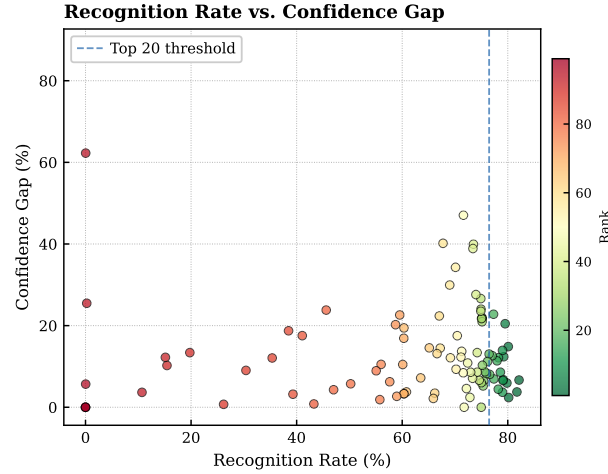
**Fig. 4.** Recognition Rate achieved by all teams with valid submissions in the Blind Test Phase, sorted by final rank. The dashed horizontal line indicates the mean Recognition Rate across all participants, while the top 20 teams are highlighted in blue. Six teams (ranked 94<sup>th</sup> to 99<sup>th</sup>) achieved Recognition Rates at or near 0% and are therefore not visually distinguishable at this scale.

Fig. 5 provides a complementary perspective by jointly analyzing Recognition Rate and Confidence Gap. The scatter plot shows that methods with similar recognition performance can differ substantially in Confidence Gap. For instance, the 3<sup>rd</sup>-ranked team achieved 80.17% Recognition Rate with a Confidence Gap of only 2.38%, whereas the 4<sup>th</sup>- and 6<sup>th</sup>-ranked teams obtained slightly lower Recognition Rates, namely 80.10% and 79.50%, but much higher Confidence Gaps of 14.86% and 20.47%, respectively. A particularly illustrative example is the 17<sup>th</sup>-ranked team, which achieved the highest Confidence Gap among the top 20, namely 22.80%, despite a Recognition Rate of 77.30%.

The following section discusses the top-ranked approaches in greater detail.

### 3.1 Top-Ranked Approaches

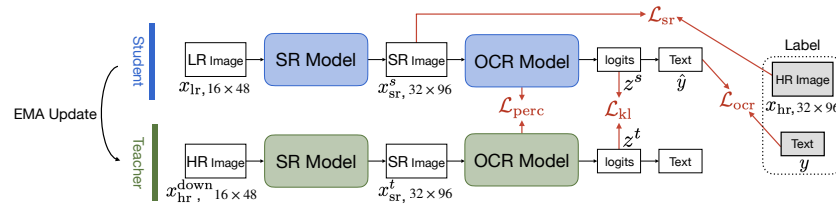
In this section, we briefly summarize the approaches of the top-5 teams in the competition. The proposed methods differ substantially in terms of architec-



**Fig. 5.** Recognition Rate versus Confidence Gap for all teams with valid submissions in the Blind Test Phase. The dashed vertical line indicates the Recognition Rate required to enter the top 20, and colors encode the final rank of each team.

ture, use of multiple frames, validation protocol, and reliance on external public datasets.

**1<sup>st</sup> Place** The first-place team (*DLmath*) from Korea University proposed a teacher-student framework that jointly trains a super-resolution model and an OCR model, as illustrated in Fig. 6. Specifically, the student branch takes LR inputs, whereas the teacher branch receives downsampled HR images; the teacher is initialized with weights pretrained on HR images only and is subsequently updated via Exponential Moving Average (EMA) of the student parameters. Their loss combines reconstruction, recognition, perceptual, and KL-divergence terms.



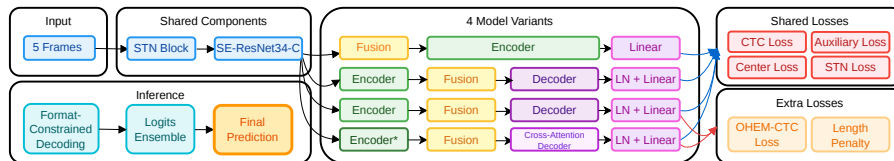
**Fig. 6.** Overview of the teacher-student framework proposed by the winning team (*DLmath*). The student branch processes low-resolution (LR) inputs while the teacher branch, updated via Exponential Moving Average (EMA), receives downsampled high-resolution (HR) images to guide the student’s learning.

HATFIR [5, 69] and MambaIRv2 [17] were adopted as super-resolution backbones, while GP-LPR [36] and a custom Transformer-based model were used for OCR. During inference, the logits from all five LR frames within each track are summed before decoding. The team further ensembled three independently trained models in the final submission by averaging their logits.

To improve robustness to partial occlusions, the team applied vertical and horizontal masking augmentations during training. From the official training data, 19,000 tracks were used for training, comprising 10,000 from Scenario-A and 9,000 from Scenario-B, while 1,000 additional Scenario-B samples were reserved for validation. The training set was further expanded with three public datasets: OpenALPR-BR [48], RodoSol-ALPR [28], and UFPR-ALPR [31].

**2<sup>nd</sup> Place** The second-place team (*AIO\_JiangnamCoffee*), composed of members from three Vietnamese institutions, namely the University of Information Technology, Ho Chi Minh University of Technology, and Vietnam National University, followed the four-stage framework of Baek et al. [2]. Their pipeline consists of a Spatial Transformer Network (STN) [24] for alignment, an SE-ResNet-C [19, 21] backbone for feature extraction, a Transformer encoder [61] for sequence modeling, and a Connectionist Temporal Classification (CTC) [16] head for prediction. A CNN-based attention module estimates the quality of each frame and fuses the five input frames into a weighted representation.

Their final system ensembles four model variants that differ in the stage at which multi-frame fusion is performed and in their use of optional Transformer decoder components, as illustrated in Fig. 7. The training objective combines primary and auxiliary CTC losses, STN loss, center loss [47], and a length penalty, with Online Hard Example Mining (OHEM) [55] included in two of the models. During inference, predictions are combined using weighted log-probability averaging, voting, and a single-model fallback strategy, while enforcing constraints associated with the Brazilian and Mercosur LP layouts.



**Fig. 7.** Overall ensemble architecture of the 2<sup>nd</sup>-place team (*AIO\_JiangnamCoffee*). Five input frames are processed through a shared STN and SE-ResNet34 backbone and fed into four model variants that differ in Transformer encoder depth and the use of auxiliary losses (OHEM-CTC and length penalty).

The models were trained from scratch for 30 epochs using AdamW [38] and a one-cycle learning rate schedule. To effectively double the training set, HR frames were synthetically degraded using blur, noise, compression, and down-

scaling before standard augmentations such as affine and perspective transformations, color shifts, and coarse dropout were applied. Only the official training data were used. The validation set was drawn exclusively from Scenario-B, which is the scenario used in the test data, following a 90/10 train-validation split.

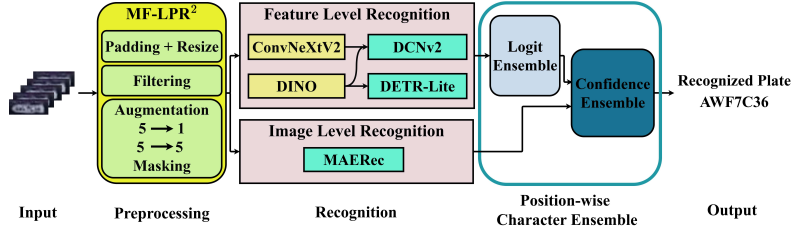
**3<sup>rd</sup> Place** The third-place team (*OpenOCR*), from the Institute of Trustworthy Embodied AI at Fudan University and the Shanghai Key Laboratory of Multimodal Embodied AI, both in China, treated LRLPR as a robust scene text recognition problem. Rather than applying a dedicated super-resolution stage, each LR frame was directly fed to an OCR model, and the five per-frame predictions within a track were aggregated using a character-level voting strategy that combines both prediction frequency and per-character confidence scores.

After comparing several candidate OCR models, including those proposed in [3, 11, 54], the team adopted SVTRv2-AR [10, 68] as their backbone, an attention-based autoregressive model whose visual encoder combines CNN and Transformer components. Four SVTRv2-AR models were trained using two dataset configurations (official training data only vs. official training data augmented with RodoSol-ALPR [28] and UFPR-ALPR [31]) and two initialization strategies (with and without Union14M-L [25] and TextSSR [67] pretraining). During inference, each track yields 20 predictions (5 frames  $\times$  4 models), which are fused using character-level majority voting, with ties resolved by confidence score.

Training used AdamW [38] with a one-cycle learning rate schedule over 100 epochs, including 10 warm-up epochs. Data augmentation followed the PARSeq [3] protocol. A small validation set of 25 tracks (125 images) was held out from the official training data to monitor overfitting. The model contains 22.42M parameters and was trained on 8 NVIDIA V100 GPUs.

**4<sup>th</sup> Place** The fourth-place team (*Capture And Predict Plate – CAP2*), from Handong Global University (Republic of Korea), proposed a multi-stage pipeline combining geometry-aware preprocessing, dual-stream recognition, and position-wise ensemble, as illustrated in Fig. 8. The preprocessing stage extends MF-LPR2 [42] with padding, resizing, filtering, and background suppression via U-Net [52]-generated text-region masks. A key modification relative to the original MF-LPR2 is that, instead of fusing five frames into a single restored image (5  $\rightarrow$  1), each frame is treated as an independent restored candidate (5  $\rightarrow$  5).

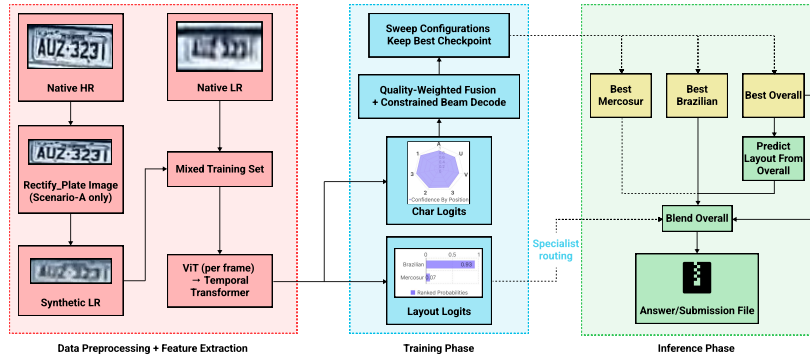
Recognition combined multiple feature extractors (ConvNeXtV2 [65], DINOv2 [49], DINOv3 [57]) and recognizers (DETR-Lite [4], DCNv2 [70]) at the feature level, alongside MAERec-B [25] at the image level. Test-time augmentation averaged logits across brightness, contrast, and sharpness variants. A two-stage position-wise character ensemble was then applied: first, a logit-level ensemble among the feature-level models using Optuna-optimized weights, followed by confidence-based fusion with the image-level predictions.



**Fig. 8.** Overview of the pipeline proposed by the 4<sup>th</sup>-place team (*CAP2*). It comprises MF-LPR<sup>2</sup>-based preprocessing, dual-stream recognition with feature-level and image-level branches, and a two-stage position-wise character ensemble for final prediction.

All models were trained exclusively on the official training data, with 10% held out for validation, using AdamW [38] with *CosineWarmRestart* or *CosineAnnealing* scheduling. No external or synthetic data were used.

**5<sup>th</sup> Place** The fifth-place team (*UIT-MeoBeo*), from the University of Information Technology and Vietnam National University, both in Vietnam, proposed a multi-stage, multi-frame OCR pipeline combining geometry-aware preprocessing, Transformer-based recognition, and structure-aware decoding (see Fig. 9). The recognition backbone uses a PE-Core-L-14-336 frame encoder [22] and a two-layer temporal Transformer for cross-frame fusion, with dual prediction heads jointly estimating the LP text and the layout (Brazilian or Mercosur).



**Fig. 9.** Overview of the pipeline proposed by the 5<sup>th</sup>-place team (*UIT-MeoBeo*). Native LR images and synthetic LR samples generated from HR license plates were used to train a ViT-based multi-frame recognizer with dual heads for character and layout estimation. During inference, quality-weighted constrained decoding and layout-specific specialist routing were used to obtain the final prediction.

Key design choices include: (i) structure-aware decoding with fixed 7-character output and position-wise letter/digit constraints; (ii) layout-aware dual

decoding, in which both Brazilian and Mercosur character masks are applied when the LP layout is uncertain, with the best hypothesis selected according to beam score and layout prior; (iii) quality-aware multi-frame fusion using sharpness and noise proxies; and (iv) a specialist cascade that routes predictions to layout-specific models when confidence is high. Training mixed native LR images with synthetic LR samples generated from HR images via downscaling, upscaling, blur, JPEG compression, noise, and sharpening.

Training used AdamW [38] with a cosine decay schedule and warmup [37], incorporating gradient clipping, mixed precision [40], EMA of model parameters, and early stopping. Weighted sampling and layout-specific loss weighting were used to maintain class balance. The model contains approximately 300 million parameters and was trained on an NVIDIA RTX 4090 GPU.

**Approaches Ranked from 6<sup>th</sup> to 10<sup>th</sup>** Among the teams ranked from 6<sup>th</sup> to 10<sup>th</sup>, only three submitted method summaries. Although these notes are less detailed than those provided by the top-5 teams, they still offer useful insights into additional design choices explored in the competition.

The 6<sup>th</sup>-place team adopted a recognition-oriented strategy that avoided a full super-resolution stage, instead relying on lightweight enhancement, coarse alignment with an STN [24], HRNet-based [58] multi-scale feature extraction, and a column-wise temporal fusion mechanism tailored to CTC decoding [16]. The 7<sup>th</sup>-place team proposed an ensemble of heterogeneous recognizers, including ResNet+Transformer-based, Mamba-based [17, 18], and SVTR-based [9, 10] models. The 9<sup>th</sup>-place team, in turn, introduced a pipeline combining spatial rectification, attention-based fusion of the five input frames, Transformer-based sequence modeling, and strategies such as EMA and test-time augmentation.

## 4 Discussion

The results show that the competition was highly competitive and that the task remains clearly unsolved. At the top of the leaderboard, the margins were very small: four teams exceeded 80% Recognition Rate, the top 10 were separated by only 3.13 percentage points, and the gap between the 1<sup>st</sup>- and 20<sup>th</sup>-ranked teams was only 5.66 percentage points. At the same time, the broader spread of scores and the winner’s residual error rate of 17.87% indicate that LRLPR remains challenging under the exact-match evaluation protocol. Taken together, these findings suggest that the field has already developed strong solutions, but still has considerable room for improvement.

Another important takeaway is that there was no single dominant design among the best-performing methods. Top teams adopted substantially different strategies, including explicit super-resolution, direct recognition from LR inputs, lightweight and heavy backbones, and different choices regarding the use of external public datasets. What appears more consistent across the strongest submissions is the effective use of the five-frame track structure, whether through temporal fusion, voting, logit aggregation, or ensemble-based integration. This

indicates that future gains may depend less on committing to a specific architectural family and more on how well methods exploit complementary information across frames while respecting the structure of Brazilian and Mercosur LPs.

The results also highlight *Confidence Gap* as an important complementary target for future work. As shown in Fig. 5, methods with similar Recognition Rates can differ substantially in Confidence Gap, which suggests that recognition accuracy and confidence quality capture different aspects of system behavior. This distinction is relevant in practice, as confidence scores can support manual verification, candidate prioritization in forensic workflows, rejection of uncertain outputs, and the combination of evidence across frames. Accordingly, improving LRLPR systems should involve not only increasing Recognition Rate, but also making confidence estimates more informative and better aligned with actual correctness.

## 5 Conclusions

In this paper, we presented the *ICPR 2026 Competition on Low-Resolution License Plate Recognition*, covering the dataset, evaluation protocol, participation statistics, final results, and summaries of the top-ranked approaches. To the best of our knowledge, this competition established the first large-scale international benchmark specifically focused on LRLPR with real low-quality data collected under operationally relevant conditions. The participation of 269 teams from 41 countries also demonstrates the practical importance of the problem and the strong interest of the community in robust ALPR under adverse conditions.

Beyond the final ranking itself, the competition provided a useful snapshot of the main methodological directions currently driving progress in the area. The top submissions showed that strong performance can be achieved through different combinations of restoration, recognition, temporal modeling, ensembling, and layout-aware decoding. This methodological diversity is one of the main outcomes of the benchmark, as it offers a clearer basis for future comparisons and for the design of stronger baselines.

We expect this benchmark to support the next stage of research on robust ALPR. Promising directions include improved multi-frame modeling, better confidence estimation, stronger use of layout constraints, and tighter integration between restoration and recognition. More broadly, the dataset, leaderboard, and method summaries may also stimulate follow-up studies on related problems such as LP super-resolution, LP legibility assessment, and forensic search support. We hope this initiative helps advance recognition systems that are not only more accurate, but also more reliable in real-world applications.

## Acknowledgments

This study was supported in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES) – Finance Code 001, through the *Programa*

de Excelência Acadêmica (PROEX); in part by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) under grant # 315409/2023-1; and in part by the *Fundação Araucária* under grant # 078/2026. The authors also thank the *Pontifícia Universidade Católica do Paraná* (PUCPR) for the financial support that enabled conference participation.

## References

1. Aharon, N., Orfaig, R., Bobrovsky, B.Z.: BoT-SORT: Robust associations multi-pedestrian tracking. arXiv preprint arXiv:2104.07636 pp. 1–13 (2022)
2. Baek, J., Kim, G., Lee, J., Park, S., Han, D., Yun, S., Oh, S.J., Lee, H.: What is wrong with scene text recognition model comparisons? dataset and model analysis. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4714–4722 (2019). <https://doi.org/10.1109/ICCV.2019.00481>
3. Bautista, D., Atienza, R.: Scene text recognition with permuted autoregressive sequence models. In: European Conference on Computer Vision (ECCV). pp. 178–196 (2022). [https://doi.org/10.1007/978-3-031-19815-1\\_11](https://doi.org/10.1007/978-3-031-19815-1_11)
4. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European Conference on Computer Vision (ECCV). pp. 213–229 (2020). [https://doi.org/10.1007/978-3-030-58452-8\\_13](https://doi.org/10.1007/978-3-030-58452-8_13)
5. Chen, X., Wang, X., Zhang, W., Kong, X., Qiao, Y., Zhou, J., Dong, C.: HAT: Hybrid attention transformer for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **48**(3), 2676–2694 (2026). <https://doi.org/10.1109/TPAMI.2025.3628275>
6. Chen, Z., et al.: NTIRE 2025 challenge on image super-resolution ( $\times 4$ ): Methods and results. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). pp. 1516–1526 (2025). <https://doi.org/10.1109/CVPRW67362.2025.00141>
7. Coluccia, A., Fascista, A., Dimou, A., Zarpalas, D., Sommer, L., Schumann, A., Mele, E.: The drone-vs-bird detection grand challenge at IJCNN 2025. In: International Joint Conference on Neural Networks (IJCNN). pp. 1–8 (2025). <https://doi.org/10.1109/IJCNN64981.2025.11228314>
8. Diwan, T., Anirudh, G., Tembhurne, J.V.: Object detection using YOLO: challenges, architectural successors, datasets and applications. *Multimedia Tools and Applications* **82**(6), 9243–9275 (2023). <https://doi.org/10.1007/s11042-022-13644-y>
9. Du, Y., Chen, Z., Jia, C., Yin, X., Zheng, T., Li, C., Du, Y., Jiang, Y.G.: SVTR: Scene text recognition with a single visual model. In: International Joint Conference on Artificial Intelligence (IJCAI). pp. 884–890 (2022). <https://doi.org/10.24963/ijcai.2022/124>
10. Du, Y., Chen, Z., Xie, H., Jia, C., Jiang, Y.G.: SVTRv2: CTC beats encoder-decoder models in scene text recognition. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 20147–20156 (Oct 2025)
11. Du, Y., Zhao, M., Fan, S., Chen, Z., Jia, C., Jiang, Y.G.: MDiff4STR: Mask diffusion model for scene text recognition. In: AAAI Conference on Artificial Intelligence. pp. 3705–3713 (Mar 2026). <https://doi.org/10.1609/aaai.v40i5.37370>
12. Gonçalves, G.R., Diniz, M.A., Laroça, R., Menotti, D., Schwartz, W.R.: Real-time automatic license plate recognition through deep multi-task networks. In:

- Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 110–117 (Oct 2018). <https://doi.org/10.1109/SIBGRAPI.2018.00021>
13. Gonçalves, G.R., Diniz, M.A., Laroca, R., Menotti, D., Schwartz, W.R.: Multi-task learning for low-resolution license plate recognition. In: Iberoamerican Congress on Pattern Recognition (CIARP). pp. 251–261 (Oct 2019). [https://doi.org/10.1007/978-3-030-33904-3\\_23](https://doi.org/10.1007/978-3-030-33904-3_23)
  14. Gong, H., Feng, Y., Zhang, Z., Hou, X., Liu, J., Huang, S., Liu, H.: A dataset and model for realistic license plate deblurring. In: International Joint Conference on Artificial Intelligence (IJCAI). pp. 1–9 (2024). <https://doi.org/10.24963/ijcai.2024/86>
  15. Gong, H., Zhang, Z., Feng, Y., Nguyen, A., Liu, H.: LP-Diff: Towards improved restoration of real-world degraded license plate. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 17831–17840 (2025). <https://doi.org/10.1109/CVPR52734.2025.01661>
  16. Graves, A., Fernández, S., Gomez, F., Schmidhuber, J.: Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In: International Conference on Machine Learning (ICML). pp. 369–376 (2006). <https://doi.org/10.1145/1143844.1143891>
  17. Guo, H., Guo, Y., Zha, Y., Zhang, Y., Li, W., Dai, T., Xia, S.T., Li, Y.: MambaRv2: Attentive state space restoration. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 28124–28133 (2025). <https://doi.org/10.1109/CVPR52734.2025.02619>
  18. Guo, H., Li, J., Dai, T., Ouyang, Z., Ren, X., Xia, S.T.: MambaIR: A simple baseline for image restoration with state-space model. In: European Conference on Computer Vision (ECCV). pp. 222–241 (2025). [https://doi.org/10.1007/978-3-031-72649-1\\_13](https://doi.org/10.1007/978-3-031-72649-1_13)
  19. He, T., Zhang, Z., Zhang, H., Zhang, Z., Xie, J., Li, M.: Bag of tricks for image classification with convolutional neural networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 558–567 (2019). <https://doi.org/10.1109/CVPR.2019.00065>
  20. Hsu, G.S., Chen, J.C., Chung, Y.Z.: Application-oriented license plate recognition. *IEEE Transactions on Vehicular Technology* **62**(2), 552–561 (Feb 2013). <https://doi.org/10.1109/TVT.2012.2226218>
  21. Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7132–7141 (2018). <https://doi.org/10.1109/CVPR.2018.00745>
  22. Hugging Face: Model card and pretrained weights for pe-core-l14-336. Hugging Face model hub (2025), timm-remapped image-encoder-only variant
  23. Ismail, A., Mehri, M., Sahbani, A., Essoukri Ben Amara, N.: Automatic license plate recognition in in-the-wild scenarios: A comprehensive review, open issues, and future directions. *IEEE Access* **13**, 145387–145415 (2025). <https://doi.org/10.1109/ACCESS.2025.3598971>
  24. Jaderberg, M., Simonyan, K., Zisserman, A., Kavukcuoglu, K.: Spatial transformer networks. In: International Conference on Neural Information Processing Systems (NeurIPS). pp. 2017–2025 (2015)
  25. Jiang, Q., Wang, J., Peng, D., Liu, C., Jin, L.: Revisiting scene text recognition: A data perspective. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 20486–20497 (2023). <https://doi.org/10.1109/ICCV51070.2023.01878>

26. Ke, X., Zeng, G., Guo, W.: An ultra-fast automatic license plate recognition approach for unconstrained scenarios. *IEEE Transactions on Intelligent Transportation Systems* **24**(5), 5172–5185 (2023). <https://doi.org/10.1109/TITS.2023.3237581>
27. Kim, D., Kim, J., Park, E.: AFA-Net: Adaptive feature attention network in image deblurring and super-resolution for improving license plate recognition. *Computer Vision and Image Understanding* **238**, 103879 (2024). <https://doi.org/10.1016/j.cviu.2023.103879>
28. Laroca, R., Cardoso, E.V., Lucio, D.R., Estevam, V., Menotti, D.: On the cross-dataset generalization in license plate recognition. In: *International Conference on Computer Vision Theory and Applications (VISAPP)*. pp. 166–178 (Feb 2022). <https://doi.org/10.5220/0010846800003124>
29. Laroca, R., dos Santos, M., Menotti, D.: Improving small drone detection through multi-scale processing and data augmentation. In: *International Joint Conference on Neural Networks (IJCNN)*. pp. 1–8 (June 2025). <https://doi.org/10.1109/IJCNN64981.2025.11227421>
30. Laroca, R., Estevam, V., Britto Jr., A.S., Minetto, R., Menotti, D.: Do we train on test data? The impact of near-duplicates on license plate recognition. In: *International Joint Conference on Neural Networks (IJCNN)*. pp. 1–8 (June 2023). <https://doi.org/10.1109/IJCNN54540.2023.10191584>
31. Laroca, R., Severo, E., Zanlorensi, L.A., Oliveira, L.S., Gonçalves, G.R., Schwartz, W.R., Menotti, D.: A robust real-time automatic license plate recognition based on the YOLO detector. In: *International Joint Conference on Neural Networks (IJCNN)*. pp. 1–10 (July 2018). <https://doi.org/10.1109/IJCNN.2018.8489629>
32. Laroca, R., Zanlorensi, L.A., Estevam, V., Minetto, R., Menotti, D.: Leveraging model fusion for improved license plate recognition. In: *Iberoamerican Congress on Pattern Recognition (CIARP)*. pp. 60–75 (Nov 2023). [https://doi.org/10.1007/978-3-031-49249-5\\_5](https://doi.org/10.1007/978-3-031-49249-5_5)
33. Laroca, R., Zanlorensi, L.A., Gonçalves, G.R., Todt, E., Schwartz, W.R., Menotti, D.: An efficient and layout-independent automatic license plate recognition system based on the YOLO detector. *IET Intelligent Transport Systems* **15**(4), 483–503 (2021). <https://doi.org/10.1049/itr2.12030>
34. Laroca, R., Estevam, V., Moreira, G.J.P., Minetto, R., Menotti, D.: Advancing multinational license plate recognition through synthetic and real data fusion: A comprehensive evaluation. *IET Intelligent Transport Systems* **19**(1), e70086 (2025). <https://doi.org/10.1049/itr2.70086>
35. Lima, G.E., Nascimento, V., Santos, E., Nascimento Jr., E., Laroca, R., Menotti, D.: Toward unified fine-grained vehicle classification and automatic license plate recognition. *Journal of the Brazilian Computer Society* **32**(1), 783–799 (2026). <https://doi.org/10.5753/jbcs.2026.5899>
36. Liu, Y.Y., Liu, Q., Chen, F., Yin, X.C.: Irregular license plate recognition via global information integration. In: *International Conference on Multimedia Modeling*. pp. 325–339 (2024). [https://doi.org/10.1007/978-3-031-53308-2\\_24](https://doi.org/10.1007/978-3-031-53308-2_24)
37. Loshchilov, I., Hutter, F.: SGDR: Stochastic gradient descent with warm restarts. In: *International Conference on Learning Representations (ICLR)*. pp. 1–16 (2017)
38. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: *International Conference on Learning Representations (ICLR)*. pp. 1–19 (2019)
39. Maier, A., Moussa, D., Spruck, A., Seiler, J., Riess, C.: Reliability scoring for the recognition of degraded license plates. In: *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. pp. 1–8 (2022). <https://doi.org/10.1109/AVSS56176.2022.9959390>

40. Micikevicius, P., et al.: Mixed precision training. In: International Conference on Learning Representations (ICLR). pp. 1–12 (2018)
41. Moussa, D., Maier, A., Spruck, A., Seiler, J., Riess, C.: Forensic license plate recognition with compression-informed transformers. In: IEEE International Conference on Image Processing (ICIP). pp. 406–410 (Oct 2022). <https://doi.org/10.1109/ICIP46576.2022.9897178>
42. Na, K., Oh, J., Cho, Y., Kim, B., Cho, S., Choi, J., Kim, I.: MF-LPR2: Multi-frame license plate image restoration and recognition using optical flow. *Computer Vision and Image Understanding* **256**, 104361 (2025). <https://doi.org/10.1016/j.cviu.2025.104361>
43. Nascimento, V., Laroca, R., Lambert, J.A., Schwartz, W.R., Menotti, D.: Combining attention module and pixel shuffle for license plate super-resolution. In: Conference on Graphics, Patterns and Images (SIBGRAPI). pp. 228–233 (Oct 2022). <https://doi.org/10.1109/SIBGRAPI55357.2022.9991753>
44. Nascimento, V., Laroca, R., Ribeiro, R.O., Schwartz, W.R., Menotti, D.: Enhancing license plate super-resolution: A layout-aware and character-driven approach. *Conference on Graphics, Patterns and Images (SIBGRAPI)* pp. 1–6 (2024). <https://doi.org/10.1109/SIBGRAPI62404.2024.10716303>
45. Nascimento, V., Lima, G.E., Ribeiro, R.O., Schwartz, W.R., Laroca, R., Menotti, D.: Toward advancing license plate super-resolution in real-world scenarios: A dataset and benchmark. *Journal of the Brazilian Computer Society* **1**(31), 435–449 (2025). <https://doi.org/10.5753/jbcs.2025.5159>
46. Oliveira, I.O., Laroca, R., Menotti, D., Fonseca, K.V.O., Minetto, R.: Vehicle-Rear: A new dataset to explore feature fusion for vehicle identification using convolutional neural networks. *IEEE Access* **9**, 101065–101077 (2021). <https://doi.org/10.1109/ACCESS.2021.3097964>
47. Olpadkar, K., Gavas, E.: Center loss regularization for continual learning. arXiv preprint arXiv:2110.11314 (2021)
48. OpenALPR: OpenALPR-BR dataset. <https://github.com/openalpr/benchmarks/tree/master/endoend/br> (2016)
49. Oquab, M., et al.: DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research* (2024), <https://openreview.net/forum?id=a68SUt6zFt>
50. Pan, Y., Tang, J., Tjahjadi, T.: LPSRGAN: Generative adversarial networks for super-resolution of license plate image. *Neurocomputing* **580**, 127426 (2024). <https://doi.org/10.1016/j.neucom.2024.127426>
51. Rao, Z., Yang, D., Chen, N., Liu, J.: License plate recognition system in unconstrained scenes via a new image correction scheme and improved CRNN. *Expert Systems with Applications* **243**, 122878 (2024). <https://doi.org/10.1016/j.eswa.2023.122878>
52. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI). pp. 234–241 (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
53. Schirmacher, F., Lorch, B., Maier, A., Riess, C.: Benchmarking probabilistic deep learning methods for license plate recognition. *IEEE Transactions on Intelligent Transportation Systems* **24**(9), 9203–9216 (2023). <https://doi.org/10.1109/TITS.2023.3278533>
54. Shi, B., Bai, X., Yao, C.: An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. *IEEE Transac-*

- tions on Pattern Analysis and Machine Intelligence **39**(11), 2298–2304 (Nov 2017). <https://doi.org/10.1109/TPAMI.2016.2646371>
55. Shrivastava, A., Gupta, A., Girshick, R.: Training region-based object detectors with online hard example mining. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 761–769 (2016). <https://doi.org/10.1109/CVPR.2016.89>
  56. Silva, S.M., Jung, C.R.: A flexible approach for automatic license plate recognition in unconstrained scenarios. IEEE Transactions on Intelligent Transportation Systems **23**(6), 5693–5703 (2022). <https://doi.org/10.1109/TITS.2021.3055946>
  57. Siméoni, O., et al.: DINOv3. arXiv preprint arXiv:2508.10104 pp. 1–67 (2025)
  58. Sun, K., Xiao, B., Liu, D., Wang, J.: Deep high-resolution representation learning for human pose estimation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 5686–5696 (2019). <https://doi.org/10.1109/CVPR.2019.00584>
  59. Terven, J., Córdova-Esparza, D.M., Romero-González, J.A.: A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction **5**(4), 1680–1716 (2023)
  60. Ultralytics: YOLOv11 (2026), <https://docs.ultralytics.com/models/yolo11/>, accessed: 2026-03-30
  61. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. In: International Conference on Neural Information Processing Systems (NeurIPS). p. 6000–6010 (2017)
  62. Wei, C., Han, F., Fan, Z., Shi, L., Peng, C.: Efficient license plate recognition in unconstrained scenarios. Journal of Visual Communication and Image Representation **104**, 104314 (2024). <https://doi.org/10.1016/j.jvcir.2024.104314>
  63. Weihong, W., Jiaoyang, T.: Research on license plate recognition algorithms based on deep learning in complex environment. IEEE Access **8**, 91661–91675 (2020). <https://doi.org/10.1109/ACCESS.2020.2994287>
  64. Wojcik, L., Machoski, E.A.F., Nascimento Jr., E., Laroca, R., Menotti, D.: LPLCv2: An expanded dataset for fine-grained license plate legibility classification. International Joint Conference on Neural Networks (IJCNN) pp. 1–7 (2026)
  65. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I.S., Xie, S.: ConvNeXt V2: Co-designing and scaling convnets with masked autoencoders. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 16133–16142 (2023). <https://doi.org/10.1109/CVPR52729.2023.01548>
  66. Xu, Z., Escalera, S., Pavão, A., Richard, M., Tu, W.W., Yao, Q., Zhao, H., Guyon, I.: Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. Patterns **3**(7), 100543 (2022). <https://doi.org/10.1016/j.patter.2022.100543>
  67. Ye, X., Du, Y., Tao, Y., Chen, Z.: TextSSR: Diffusion-based data synthesis for scene text recognition. In: IEEE/CVF International Conference on Computer Vision (ICCV). pp. 17464–17473 (2025)
  68. Ye, X., Du, Y., Zhang, J., Li, C., LYU, J., Chen, Z.: What’s wrong with synthetic data for scene text recognition? a strong synthetic engine with diverse simulations and self-evolution. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2026)
  69. Zhang, D., Huang, F., Liu, S., Wang, X., Jin, Z.: SwinFIR: Revisiting the SwinIR with fast fourier convolution and improved training for image super-resolution. arXiv preprint arXiv:2208.11247 pp. 1–14 (2023)
  70. Zhu, X., Hu, H., Lin, S., Dai, J.: Deformable convnets v2: More deformable, better results. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9300–9308 (2019). <https://doi.org/10.1109/CVPR.2019.00953>