Automatic License Plate Recognition: Toward Improving the State of the Art and Bridging the Gap Between Academia and Industry

Rayson Laroca

Advisor: David Menotti
Co-advisor: Rodrigo Minetto

PhD Final Examination

Thesis Outline

1. Introduction;
2. Theoretical Foundation;
3. Related Work;
4. The RodoSol-ALPR Dataset;
5. Cross-Dataset Generalization;
6. Model Fusion;
7. Synthetic Data;
8. Near-Duplicates;
9. Dataset Bias;
10. Conclusions.

Thesis Outline

Automatic License Plate Recognition (ALPR)

A typical *Automatic License Plate Recognition (ALPR)* system.

Automatic License Plate Recognition (ALPR)

A typical *Automatic License Plate Recognition (ALPR)* system.

ALPR has many practical applications:

- Toll collection;
- Vehicle access control in restricted areas;
- Traffic law enforcement.

Automatic License Plate Recognition (ALPR)

A typical *Automatic License Plate Recognition (ALPR)* system.

ALPR has many practical applications:

- Toll collection;
- Vehicle access control in restricted areas;
- Traffic law enforcement.

Current research has mostly focused on the **License Plate Recognition (LPR)** stage.

Problem Statement – Internationalization

ALPR systems must handle LPs from different regions with different character sets.

Examples of different LP styles in the United States.

Problem Statement – Internationalization

ALPR systems must handle LPs from different regions with different character sets.

Examples of different LP styles in the United States.

Most ALPR systems presented in the literature were designed specifically for **a single LP style** (e.g., single-row blue LPs from mainland China).

Problem Statement – Mercosur LPs

Mercosur[1] countries have adopted a unified standard of LPs for newly purchased vehicles.

[1]Mercosur, short for *Mercado Común del Sur* (Southern Common Market in Spanish), is an economic and political bloc comprising Argentina, Brazil, Paraguay and Uruguay.

Problem Statement – Mercosur LPs

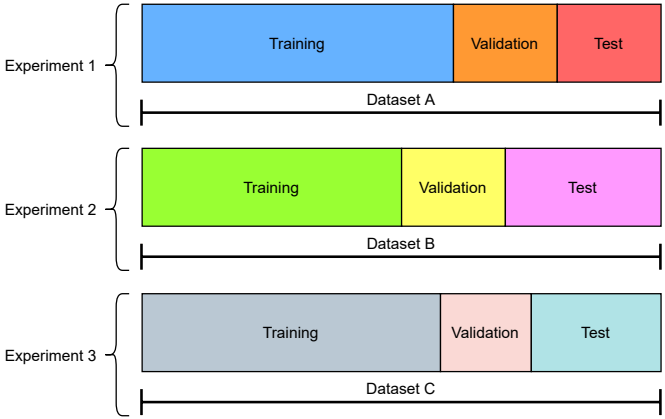
Mercosur[1] countries have adopted a unified standard of LPs for newly purchased vehicles.

A public dataset containing images of Mercosur LPs **is not yet available**!

[1] Mercosur, short for *Mercado Común del Sur* (Southern Common Market in Spanish), is an economic and political bloc comprising Argentina, Brazil, Paraguay and Uruguay.

Problem Statement – Evaluation Protocols [1/3]

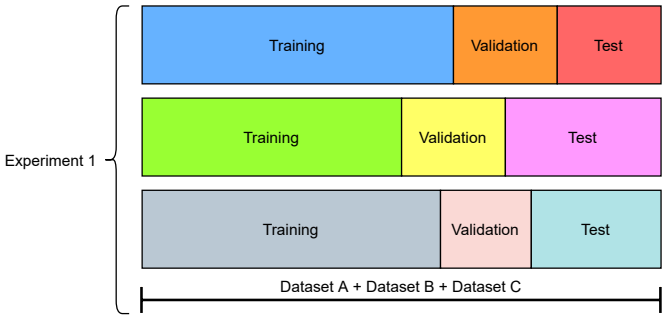
In the past, the evaluation of ALPR systems used to be done within individual datasets.



The proposed methods were trained/adjusted **multiple times**, once for each dataset.

Problem Statement – Evaluation Protocols [2/3]

Recently, the proposed models have been **trained once** on the union of the training images from the selected datasets and evaluated separately on the respective test sets.



- ALPR systems have often achieved **impressive results** under this protocol.

Problem Statement – Evaluation Protocols [3/3]

Generalization ability?

Problem Statement – Evaluation Protocols [3/3]

Generalization ability?

- In real-world applications, **new cameras are regularly being installed in new locations** without existing ALPR systems being retrained as often.

Problem Statement – Labeled Data & Privacy Concerns

- **Labeled data is expensive:**
  - "*Real data is not easy to obtain, the acquisition process is slow, and the data needs to be processed and annotated before it can be used for training. To achieve a higher accuracy of the annotation, manual inspection is also required.*" — Wu et al. (2018);
  - "*Collecting a sufficient number of LP images is extremely difficult for normal research.*" — Han et al. (2020);
  - "*Reducing the number of human-labeled samples or interactions with the world that are required to learn a task is of crucial importance.*" — Bengio et al. (2021).

Problem Statement – Labeled Data & Privacy Concerns

- Labeled data is expensive;
- **Privacy concerns are growing:**
  - National data protection law that came into force in May 2018 in Europe
    - All personal items that can be used for identification must be anonymized in captured data
    - In Europe, license plate number can be used for identification

Problem Statement – Labeled Data & Privacy Concerns

- Labeled data is expensive;
- Privacy concerns are growing.

Q: How to reduce the number of real and human-labeled images needed to achieve the expected results?

Problem Statement – Labeled Data & Privacy Concerns

- Labeled data is expensive;
- Privacy concerns are growing.

Q: How to reduce the number of real and human-labeled images needed to achieve the expected results?

A: Synthetic Data!

Hypothesis and Research Questions

Hypothesis

It is possible to significantly improve the state of the art in ALPR without increasing the number of real training images, designing groundbreaking descriptors, or extensively searching for better model architectures.

Hypothesis and Research Questions

Hypothesis

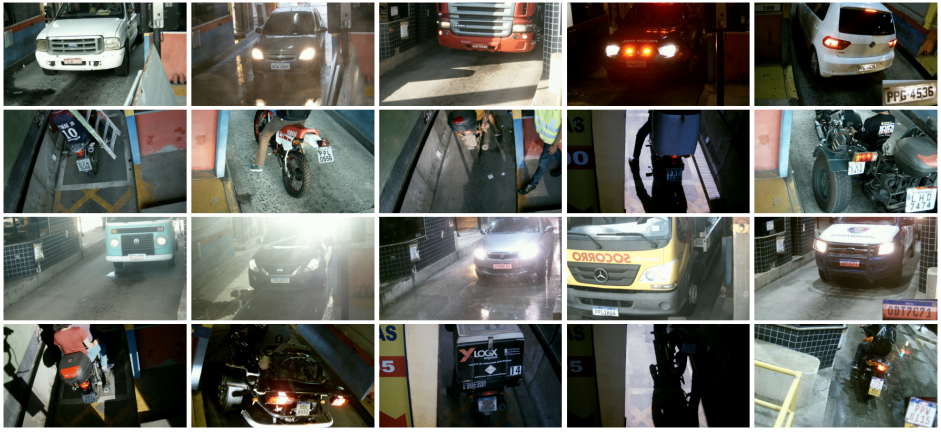
It is possible to significantly improve the state of the art in ALPR without increasing the number of real training images, designing groundbreaking descriptors, or extensively searching for better model architectures.

Some questions that guide our research are:

- How can we address the lack of attention given to images featuring Mercosur LPs?

The RodoSol-ALPR Dataset

RodoSol-ALPR Dataset [1/3]



https://github.com/raysonlaroca/rodosol-alpr-dataset/

- RodoSol-ALPR contains **20,000 images** (1,280 × 720 pixels) captured by static cameras located at pay tolls owned by the *Rodovia do Sol* (RodoSol) concessionaire.

RodoSol-ALPR Dataset [2/3]

Some LPs from the RodoSol-ALPR dataset.

- 5,000 images of cars with Brazilian LPs (1st row);
- 5,000 images of motorcycles with Brazilian LPs (2nd row);
- 5,000 images of cars with Mercosur LPs (3rd row);
- 5,000 images of motorcycles with Mercosur LPs (4th row).

RodoSol-ALPR Dataset [3/3]

Access – **146 researchers** from **42 countries** around the world:

https://raysonlaroca.github.io/misc/rodosol-alpr-map/

Recap – Hypothesis and Research Questions

Hypothesis

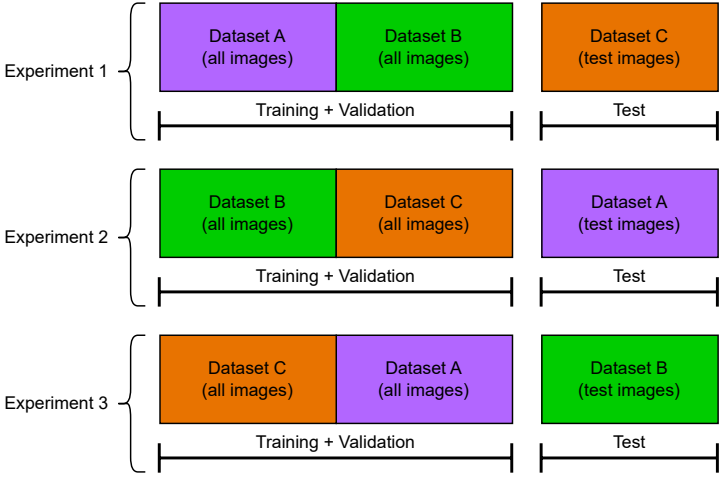
It is possible to significantly improve the state of the art in ALPR without increasing the number of real training images, designing groundbreaking descriptors, or extensively searching for better model architectures.

Some questions that guide our research are:

- How can we address the lack of attention given to images featuring Mercosur LPs?
- Do current methods for detecting and recognizing LPs generalize well to unseen data?

On the Cross-Dataset Generalization in License Plate Recognition

Experimental Setup – Leave-One-Dataset-Out Protocol



Experimental Setup – Overview

A **traditional-split** *versus* **leave-one-dataset-out** experimental setup:

- **12** OCR models;
- RodoSol-ALPR + **8** well-known public datasets.

Experimental Setup – OCR Models

The 12 OCR models explored in this chapter.

Model	Original Application
Framework: PyTorch	
R^2AM (Lee and Osindero, 2016)	Scene Text Recognition
RARE (Shi et al., 2016)	Scene Text Recognition
STAR-Net (Liu et al., 2016)	Scene Text Recognition
CRNN (Shi et al., 2017)	Scene Text Recognition
GRCNN (Wang and Hu, 2017)	Scene Text Recognition
Rosetta (Borisyuk et al., 2018)	Scene Text Recognition
TRBA (Baek et al., 2019)	Scene Text Recognition
ViTSTR-Base (Atienza, 2021)	Scene Text Recognition
Framework: Keras	
Holistic-CNN (Špaňhel et al., 2017)	License Plate Recognition
Multi-Task-LR (Gonçalves et al., 2019)	License Plate Recognition
Framework: Darknet	
CR-NET (Silva and Jung, 2020)	License Plate Recognition
Fast-OCR (Laroca et al., 2021)	Image-based Meter Reading

Experimental Setup – OCR Models

The 12 OCR models explored in this chapter.

Model	Original Application
Framework: PyTorch	
R^2AM (Lee and Osindero, 2016)	Scene Text Recognition
RARE (Shi et al., 2016)	Scene Text Recognition
STAR-Net (Liu et al., 2016)	Scene Text Recognition
CRNN (Shi et al., 2017)	Scene Text Recognition
GRCNN (Wang and Hu, 2017)	Scene Text Recognition
Rosetta (Borisyuk et al., 2018)	Scene Text Recognition
TRBA (Baek et al., 2019)	Scene Text Recognition
ViTSTR-Base (Atienza, 2021)	Scene Text Recognition
Framework: Keras	
Holistic-CNN (Špaňhel et al., 2017)	License Plate Recognition
Multi-Task-LR (Gonçalves et al., 2019)	License Plate Recognition
Framework: Darknet	
CR-NET (Silva and Jung, 2020)	License Plate Recognition
Fast-OCR (Laroca et al., 2021)	Image-based Meter Reading

Experimental Setup – Datasets [1/2]

RodoSol-ALPR + 8 public datasets:

(a) Caltech Cars

(b) EnglishLP

(c) UCSD-Stills

(d) ChineseLP

(e) AOLP

(f) OpenALPR-EU

(g) SSIG-SegPlate

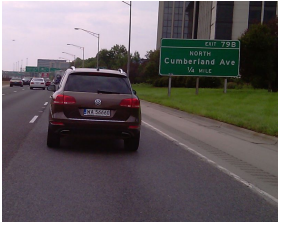
(h) UFPR-ALPR

Experimental Setup – Datasets [2/2]

Original images:







Results – LP Detection

Recall rates obtained by YOLOv4 in the **LP detection** stage (IoU ≥ 0.5).

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	OpenALPR-EU # 108	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
Traditional-split	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	99.9%	99.1%	100.0%	**99.9%**
Leave-one-dataset-out	100.0%	100.0%	100.0%	100.0%	99.9%	99.1%	100.0%	96.8%	99.6%	**99.5%**

Results – LP Detection

Recall rates obtained by YOLOv4 in the **LP detection** stage (IoU ≥ 0.5).

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	OpenALPR-EU # 108	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
Traditional-split	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	99.9%	99.1%	100.0%	**99.9%**
Leave-one-dataset-out	100.0%	100.0%	100.0%	100.0%	99.9%	99.1%	100.0%	96.8%	99.6%	**99.5%**

Recall rates above **99.9%** were achieved in 14 of the 18 assessments.

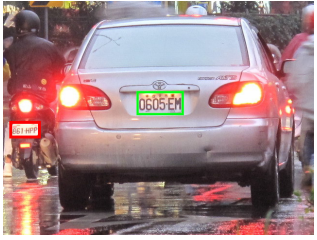
Results – LP Detection

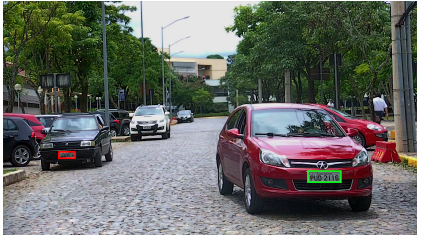
Recall rates obtained by YOLOv4 in the **LP detection** stage (IoU ≥ 0.5).

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	OpenALPR-EU # 108	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
Traditional-split	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	99.9%	99.1%	100.0%	**99.9%**
Leave-one-dataset-out	100.0%	100.0%	100.0%	100.0%	99.9%	99.1%	100.0%	96.8%	99.6%	**99.5%**

Recall rates above **99.9%** were achieved in 14 of the 18 assessments.

- Regarding the precision rates, the "false positives" identified by YOLOv4 primarily correspond to unlabeled LPs in the image backgrounds, not actual errors:





Results – LP Recognition (Traditional-Split)

Recognition rates obtained by all models under the **traditional-split** protocol.

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	OpenALPR-EU # 108	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	94.1%	**100.0%**	**97.5%**	98.0%	96.3%	**97.5%**	82.6%	59.0%†	91.4%
CRNN	93.5%	88.2%	91.7%	90.7%	97.1%	93.5%	92.9%	68.9%	73.6%	87.8%
Fast-OCR	93.5%	**97.1%**	**100.0%**	**97.5%**	98.1%	**97.2%**	97.1%	81.6%	56.7%†	91.0%
GRCNN	93.5%	92.2%	93.3%	91.9%	97.1%	87.0%	93.4%	66.6%	77.6%	88.1%
Holistic-CNN	87.0%	75.5%	88.3%	95.0%	97.7%	89.8%	95.6%	81.2%	94.7%	89.4%
Multi-Task-LR	89.1%	73.5%	85.0%	92.5%	94.9%	85.2%	93.3%	72.3%	86.6%	85.8%
R^2AM	89.1%	83.3%	86.7%	91.9%	96.5%	88.9%	92.0%	75.9%	83.4%	87.5%
RARE	95.7%	94.1%	95.0%	94.4%	97.7%	94.4%	94.0%	75.7%	78.7%	91.1%
Rosetta	89.1%	82.4%	93.3%	93.8%	97.5%	90.7%	94.4%	75.5%	89.0%	89.5%
STAR-Net	95.7%	96.1%	95.0%	95.7%	97.8%	**97.2%**	96.1%	78.8%	82.3%	**92.7%**
TRBA	93.5%	91.2%	91.7%	93.8%	97.2%	93.5%	97.3%	83.4%	80.6%	91.3%
ViTSTR-Base	87.0%	88.2%	86.7%	96.9%	**99.4%**	89.8%	95.8%	**89.7%**	**95.6%**	92.1%
Average	92.0%	88.0%	92.2%	94.3%	97.4%	92.0%	95.0%	77.7%	79.8%	89.8%

Results – LP Recognition (Traditional-Split)

Recognition rates obtained by all models under the **traditional-split** protocol.

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	OpenALPR-EU # 108	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	94.1%	**100.0%**	**97.5%**	98.0%	96.3%	**97.5%**	82.6%	59.0%†	91.4%
CRNN	93.5%	88.2%	91.7%	90.7%	97.1%	93.5%	92.9%	68.9%	73.6%	87.8%
Fast-OCR	93.5%	**97.1%**	**100.0%**	**97.5%**	98.1%	**97.2%**	97.1%	81.6%	56.7%†	91.0%
GRCNN	93.5%	92.2%	93.3%	91.9%	97.1%	87.0%	93.4%	66.6%	77.6%	88.1%
Holistic-CNN	87.0%	75.5%	88.3%	95.0%	97.7%	89.8%	95.6%	81.2%	94.7%	89.4%
Multi-Task-LR	89.1%	73.5%	85.0%	92.5%	94.9%	85.2%	93.3%	72.3%	86.6%	85.8%
R^2AM	89.1%	83.3%	86.7%	91.9%	96.5%	88.9%	92.0%	75.9%	83.4%	87.5%
RARE	95.7%	94.1%	95.0%	94.4%	97.7%	94.4%	94.0%	75.7%	78.7%	91.1%
Rosetta	89.1%	82.4%	93.3%	93.8%	97.5%	90.7%	94.4%	75.5%	89.0%	89.5%
STAR-Net	95.7%	96.1%	95.0%	95.7%	97.8%	**97.2%**	96.1%	78.8%	82.3%	**92.7%**
TRBA	93.5%	91.2%	91.7%	93.8%	97.2%	93.5%	97.3%	83.4%	80.6%	91.3%
ViTSTR-Base	87.0%	88.2%	86.7%	96.9%	**99.4%**	89.8%	95.8%	**89.7%**	**95.6%**	92.1%
Average	92.0%	88.0%	92.2%	94.3%	97.4%	92.0%	95.0%	77.7%	79.8%	89.8%

Different models yield the best results on different datasets!

Results – LP Recognition (Traditional-Split)

Recognition rates obtained by all models under the **traditional-split** protocol.

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	OpenALPR-EU # 108	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	94.1%	**100.0%**	**97.5%**	98.0%	96.3%	**97.5%**	82.6%	59.0%†	91.4%
CRNN	93.5%	88.2%	91.7%	90.7%	97.1%	93.5%	92.9%	68.9%	73.6%	87.8%
Fast-OCR	93.5%	**97.1%**	**100.0%**	**97.5%**	98.1%	**97.2%**	97.1%	81.6%	56.7%†	91.0%
GRCNN	93.5%	92.2%	93.3%	91.9%	97.1%	87.0%	93.4%	66.6%	77.6%	88.1%
Holistic-CNN	87.0%	75.5%	88.3%	95.0%	97.7%	89.8%	95.6%	81.2%	94.7%	89.4%
Multi-Task-LR	89.1%	73.5%	85.0%	92.5%	94.9%	85.2%	93.3%	72.3%	86.6%	85.8%
R^2AM	89.1%	83.3%	86.7%	91.9%	96.5%	88.9%	92.0%	75.9%	83.4%	87.5%
RARE	95.7%	94.1%	95.0%	94.4%	97.7%	94.4%	94.0%	75.7%	78.7%	91.1%
Rosetta	89.1%	82.4%	93.3%	93.8%	97.5%	90.7%	94.4%	75.5%	89.0%	89.5%
STAR-Net	95.7%	96.1%	95.0%	95.7%	97.8%	**97.2%**	96.1%	78.8%	82.3%	**92.7%**
TRBA	93.5%	91.2%	91.7%	93.8%	97.2%	93.5%	97.3%	83.4%	80.6%	91.3%
ViTSTR-Base	87.0%	88.2%	86.7%	96.9%	**99.4%**	89.8%	95.8%	**89.7%**	**95.6%**	92.1%
Average	92.0%	**88.0%**	92.2%	94.3%	97.4%	92.0%	95.0%	**77.7%**	**79.8%**	89.8%

What do these datasets have in common?

Results – LP Recognition (Traditional-Split)

LPs with <u>two rows</u> of characters!

(a) EnglishLP

(b) UFPR-ALPR

(c) RodoSol-ALPR

- In Brazil, the motorcycle fleet currently represents 28% of the total vehicle fleet.[2]
  - All motorcycles in Brazil have two-row LPs.

[2] www.gov.br/infraestrutura/pt-br/assuntos/transito/conteudo-denatran/frota-de-veiculos-2024

Results – LP Recognition (Leave-One-Dataset-Out)

Recognition rates obtained by all models under the **leave-one-dataset-out** protocol.

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	OpenALPR-EU # 108	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	**97.1%**	**98.3%**	94.4%	**89.1%**	**98.1%**	97.1%	66.4%	**63.8%**	**89.1%**
CRNN	93.5%	82.4%	86.7%	84.5%	71.6%	94.4%	90.8%	62.9%	39.2%	78.4%
Fast-OCR	95.7%	95.1%	96.7%	93.8%	79.3%	96.3%	95.5%	65.9%	63.4%	86.8%
GRCNN	93.5%	82.4%	93.3%	85.1%	72.1%	91.7%	90.8%	62.7%	40.0%	79.0%
Holistic-CNN	84.8%	56.9%	76.7%	82.6%	60.0%	93.5%	93.2%	66.4%	34.5%	72.0%
Multi-Task-LR	84.8%	57.8%	78.3%	76.4%	67.5%	88.9%	90.8%	61.7%	25.2%	70.2%
R^2AM	89.1%	58.8%	81.7%	85.1%	62.6%	89.8%	94.2%	61.2%	41.1%	73.7%
RARE	89.1%	64.7%	93.3%	88.2%	70.7%	92.6%	93.9%	78.2%	40.2%	79.0%
Rosetta	95.7%	82.4%	88.3%	87.6%	70.6%	90.7%	93.9%	69.2%	42.8%	80.1%
STAR-Net	91.3%	85.3%	93.3%	92.5%	79.2%	96.3%	93.8%	74.8%	43.8%	83.4%
TRBA	91.3%	62.7%	95.0%	92.5%	75.3%	92.6%	96.8%	82.9%	42.9%	81.3%
ViTSTR-Base	93.5%	62.7%	86.7%	**96.3%**	68.9%	91.7%	**97.8%**	**84.7%**	59.7%	82.4%
Average	91.7%	74.0%	89.0%	88.3%	72.2%	93.1%	94.0%	69.7%	44.7%	79.6%
Average (traditional split)	92.0%	88.0%	92.2%	94.3%	97.4%	92.0%‡	95.0%	77.7%	79.8%	89.8%
Sighthound	87.0%	94.1%	90.0%	84.5%	79.6%	94.4%	79.2%	52.6%	51.0%	79.2%
OpenALPR	95.7%	99.0%	96.7%	93.8%	81.1%	99.1%	91.4%	87.8%	70.0%	90.5%

‡Under the traditional-split protocol, no images from the OpenALPR-EU dataset were used for training. This is the protocol commonly adopted in the literature.

Results – LP Recognition (Leave-One-Dataset-Out)

Recognition rates obtained by all models under the **leave-one-dataset-out** protocol.

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	OpenALPR-EU # 108	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	**97.1%**	**98.3%**	94.4%	**89.1%**	**98.1%**	97.1%	66.4%	**63.8%**	**89.1%**
CRNN	93.5%	82.4%	86.7%	84.5%	71.6%	94.4%	90.8%	62.9%	39.2%	78.4%
Fast-OCR	95.7%	95.1%	96.7%	93.8%	79.3%	96.3%	95.5%	65.9%	63.4%	86.8%
GRCNN	93.5%	82.4%	93.3%	85.1%	72.1%	91.7%	90.8%	62.7%	40.0%	79.0%
Holistic-CNN	84.8%	56.9%	76.7%	82.6%	60.0%	93.5%	93.2%	66.4%	34.5%	72.0%
Multi-Task-LR	84.8%	57.8%	78.3%	76.4%	67.5%	88.9%	90.8%	61.7%	25.2%	70.2%
R^2AM	89.1%	58.8%	81.7%	85.1%	62.6%	89.8%	94.2%	61.2%	41.1%	73.7%
RARE	89.1%	64.7%	93.3%	88.2%	70.7%	92.6%	93.9%	78.2%	40.2%	79.0%
Rosetta	95.7%	82.4%	88.3%	87.6%	70.6%	90.7%	93.9%	69.2%	42.8%	80.1%
STAR-Net	91.3%	85.3%	93.3%	92.5%	79.2%	96.3%	93.8%	74.8%	43.8%	83.4%
TRBA	91.3%	62.7%	95.0%	92.5%	75.3%	92.6%	96.8%	82.9%	42.9%	81.3%
ViTSTR-Base	93.5%	62.7%	86.7%	**96.3%**	68.9%	91.7%	**97.8%**	**84.7%**	59.7%	82.4%
Average	91.7%	**74.0%**	89.0%	88.3%	**72.2%**	93.1%	94.0%	69.7%	**44.7%**	79.6%
Average (traditional split)	92.0%	88.0%	92.2%	94.3%	97.4%	92.0%‡	95.0%	77.7%	79.8%	89.8%
Sighthound	87.0%	94.1%	90.0%	84.5%	79.6%	94.4%	79.2%	52.6%	51.0%	79.2%
OpenALPR	95.7%	99.0%	96.7%	93.8%	81.1%	99.1%	91.4%	87.8%	70.0%	90.5%

‡Under the traditional-split protocol, no images from the OpenALPR-EU dataset were used for training. This is the protocol commonly adopted in the literature.

Results – LP Recognition (Leave-One-Dataset-Out)

LODO: 8C83I3
Trad.: 8C8313

LODO: AB0416
Trad.: AR0416

LODO: PG379I
Trad.: P63791

LODO: 03250M
Trad.: 0325DM

The predictions obtained by TRBA on three images of the AOLP dataset.

LODO: CK3118R
Trad.: CK311BR

LODO: NB4071P
Trad.: MB4071P

LODO: -64097AC
Trad.: ZG4097AC

LODO: ZGQ880TM
Trad.: ZG 880TV

The predictions obtained by STAR-Net on three images of the EnglishLP dataset.

In general, the errors under the Leave-One-Dataset-Out (LODO) protocol did not occur in challenging cases (e.g., blurry or tilted images); therefore, they were probably caused by differences in the training and test images. Trad.: traditional-split protocol.

Results – LP Recognition (Leave-One-Dataset-Out)

Recognition rates obtained by all models under the **leave-one-dataset-out** protocol.

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	OpenALPR-EU # 108	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	**97.1%**	**98.3%**	94.4%	**89.1%**	**98.1%**	97.1%	66.4%	**63.8%**	**89.1%**
CRNN	93.5%	82.4%	86.7%	84.5%	71.6%	94.4%	90.8%	62.9%	39.2%	78.4%
Fast-OCR	95.7%	95.1%	96.7%	93.8%	79.3%	96.3%	95.5%	65.9%	63.4%	86.8%
GRCNN	93.5%	82.4%	93.3%	85.1%	72.1%	91.7%	90.8%	62.7%	40.0%	79.0%
Holistic-CNN	84.8%	56.9%	76.7%	82.6%	60.0%	93.5%	93.2%	66.4%	34.5%	72.0%
Multi-Task-LR	84.8%	57.8%	78.3%	76.4%	67.5%	88.9%	90.8%	61.7%	25.2%	70.2%
R^2AM	89.1%	58.8%	81.7%	85.1%	62.6%	89.8%	94.2%	61.2%	41.1%	73.7%
RARE	89.1%	64.7%	93.3%	88.2%	70.7%	92.6%	93.9%	78.2%	40.2%	79.0%
Rosetta	95.7%	82.4%	88.3%	87.6%	70.6%	90.7%	93.9%	69.2%	42.8%	80.1%
STAR-Net	91.3%	85.3%	93.3%	92.5%	79.2%	96.3%	93.8%	74.8%	43.8%	83.4%
TRBA	91.3%	62.7%	95.0%	92.5%	75.3%	92.6%	96.8%	82.9%	42.9%	81.3%
ViTSTR-Base	93.5%	62.7%	86.7%	**96.3%**	68.9%	91.7%	**97.8%**	**84.7%**	59.7%	82.4%
Average	91.7%	74.0%	89.0%	88.3%	72.2%	93.1%	94.0%	69.7%	**44.7%**	79.6%
Average (traditional split)	92.0%	88.0%	92.2%	94.3%	97.4%	92.0%‡	95.0%	77.7%	79.8%	89.8%
Sighthound	87.0%	94.1%	90.0%	84.5%	79.6%	94.4%	79.2%	52.6%	**51.0%**	79.2%
OpenALPR	95.7%	99.0%	96.7%	93.8%	81.1%	99.1%	91.4%	87.8%	**70.0%**	90.5%

‡Under the traditional-split protocol, no images from the OpenALPR-EU dataset were used for training. This is the protocol commonly adopted in the literature.

These results accentuated the importance of the RodoSol-ALPR dataset for training deep models for robust recognition of Mercosur and two-row LPs.

Highlights

- Researchers should pay more attention to **cross-dataset LP recognition**;
  - Significant drops in performance (e.g., 97.4% $\rightarrow$ 72.2%) when training and testing the recognition models in a leave-one-dataset-out fashion.

Highlights

- Researchers should pay more attention to **cross-dataset LP recognition**;
  - Significant drops in performance (e.g., 97.4% $\rightarrow$ 72.2%) when training and testing the recognition models in a leave-one-dataset-out fashion.
- **RodoSol-ALPR** proved essential for the reliable recognition of Mercosur LPs.
  - Both the models trained by us and two established commercial systems reached recognition rates below 70% on its test set under the leave-one-dataset-out protocol.

Highlights

- Researchers should pay more attention to **cross-dataset LP recognition**;
  - Significant drops in performance (e.g., 97.4% $\rightarrow$ 72.2%) when training and testing the recognition models in a leave-one-dataset-out fashion.
- **RodoSol-ALPR** proved essential for the reliable recognition of Mercosur LPs.
  - Both the models trained by us and two established commercial systems reached recognition rates below 70% on its test set under the leave-one-dataset-out protocol.
- Different OCR models yielded the best results on different datasets!

Recap – Hypothesis and Research Questions

Hypothesis

It is possible to significantly improve the state of the art in ALPR without increasing the number of real training images, designing groundbreaking descriptors, or extensively searching for better model architectures.

Some questions that guide our research are:

- How can we address the lack of attention given to images featuring Mercosur LPs?
- Do current methods for detecting and recognizing LPs generalize well to unseen data?
- Can we considerably improve results by combining the outputs of various OCR models?

Leveraging Model Fusion for Improved License Plate Recognition

Experimental Setup – Fusion Approaches

5 OCR Models

↓

```
ABC1234 (0.7)
ADE5678 (0.4)
ADF1235 (0.9)
ABC1234 (0.3)
ADH1236 (0.8)
```

Experimental Setup – Fusion Approaches

5 OCR Models

↓

ABC1234 (0.7)
ADE5678 (0.4)
ADF9012 (0.9)
ABC1234 (0.3)
ADH1236 (0.8)

Three primary fusion approaches:

1. **Highest Confidence (HC)**: ADF9012

Experimental Setup – Fusion Approaches

5 OCR Models

↓

ABC1234 (0.7)
ADE5678 (0.4)
ADF9012 (0.9)
ABC1234 (0.3)
ADH1236 (0.8)

Three primary fusion approaches:

1. Highest Confidence (HC): `ADF9012`
2. **Majority Vote (MV)**: `ABC1234`

Experimental Setup – Fusion Approaches

5 OCR Models

↓

ABC1234	(0.7)
ADE5678	(0.4)
ADF9012	(0.9)
ABC1234	(0.3)
ADH1236	(0.8)

Three primary fusion approaches:

1. Highest Confidence (HC): `ADF9012`
2. Majority Vote (MV): `ABC1234`
3. **Majority Vote by Character Position (MVCP)**: `ADC1234`

Results

Comparison of the recognition rates achieved across eight popular datasets by 12 models individually and through five different fusion strategies (intra-dataset experiments).

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	94.1%	**100.0%**	**97.5%**	98.1%	**97.5%**	82.6%	59.0%†	90.8%
CRNN	93.5%	88.2%	91.7%	90.7%	97.1%	92.9%	68.9%	73.6%	87.1%
Fast-OCR	93.5%	**97.1%**	**100.0%**	**97.5%**	98.1%	97.1%	81.6%	56.7%†	90.2%
GRCNN	93.5%	92.2%	93.3%	91.9%	97.1%	93.4%	66.6%	77.6%	88.2%
Holistic-CNN	87.0%	75.5%	88.3%	95.0%	97.7%	95.6%	81.2%	94.7%	89.4%
Multi-Task-LR	89.1%	73.5%	85.0%	92.5%	94.9%	93.3%	72.3%	86.6%	85.9%
R^2AM	89.1%	83.3%	86.7%	91.9%	96.5%	92.0%	75.9%	83.4%	87.4%
RARE	95.7%	94.1%	95.0%	94.4%	97.7%	94.0%	75.7%	78.7%	90.7%
Rosetta	89.1%	82.4%	93.3%	93.8%	97.5%	94.4%	75.5%	89.0%	89.4%
STAR-Net	95.7%	96.1%	95.0%	95.7%	97.8%	96.1%	78.8%	82.3%	92.2%
TRBA	93.5%	91.2%	91.7%	93.8%	97.2%	97.3%	83.4%	80.6%	91.1%
ViTSTR-Base	87.0%	88.2%	86.7%	96.9%	**99.4%**	95.8%	**89.7%**	**95.6%**	**92.4%**
Fusion HC (*top 6*)	**97.8%**	95.1%	96.7%	**98.1%**	99.0%	96.6%	90.9%	93.5%	96.0%
Fusion MV-BM (*top 8*)	**97.8%**	**97.1%**	**100.0%**	**98.1%**	**99.7%**	98.4%	92.7%	96.4%	97.5%
Fusion MV-HC (*top 8*)	**97.8%**	**97.1%**	**100.0%**	**98.1%**	**99.7%**	99.1%	92.3%	**96.5%**	**97.6%**
Fusion MVCP-BM (*top 9*)	95.7%	96.1%	**100.0%**	**98.1%**	99.6%	99.0%	**92.8%**	96.4%	97.2%
Fusion MVCP-HC (*top 9*)	**97.8%**	96.1%	**100.0%**	**98.1%**	99.6%	**99.3%**	92.5%	96.3%	97.5%

Results

Comparison of the recognition rates achieved across eight popular datasets by 12 models individually and through five different fusion strategies (intra-dataset experiments).

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	94.1%	**100.0%**	**97.5%**	98.1%	**97.5%**	82.6%	59.0%†	90.8%
CRNN	93.5%	88.2%	91.7%	90.7%	97.1%	92.9%	68.9%	73.6%	87.1%
Fast-OCR	93.5%	**97.1%**	**100.0%**	**97.5%**	98.1%	97.1%	81.6%	56.7%†	90.2%
GRCNN	93.5%	92.2%	93.3%	91.9%	97.1%	93.4%	66.6%	77.6%	88.2%
Holistic-CNN	87.0%	75.5%	88.3%	95.0%	97.7%	95.6%	81.2%	94.7%	89.4%
Multi-Task-LR	89.1%	73.5%	85.0%	92.5%	94.9%	93.3%	72.3%	86.6%	85.9%
R^2AM	89.1%	83.3%	86.7%	91.9%	96.5%	92.0%	75.9%	83.4%	87.4%
RARE	95.7%	94.1%	95.0%	94.4%	97.7%	94.0%	75.7%	78.7%	90.7%
Rosetta	89.1%	82.4%	93.3%	93.8%	97.5%	94.4%	75.5%	89.0%	89.4%
STAR-Net	95.7%	96.1%	95.0%	95.7%	97.8%	96.1%	78.8%	82.3%	92.2%
TRBA	93.5%	91.2%	91.7%	93.8%	97.2%	97.3%	83.4%	80.6%	91.1%
ViTSTR-Base	87.0%	88.2%	86.7%	96.9%	**99.4%**	95.8%	**89.7%**	**95.6%**	92.4%
Fusion HC (*top 6*)	**97.8%**	95.1%	96.7%	**98.1%**	99.0%	96.6%	90.9%	93.5%	96.0%
Fusion MV-BM (*top 8*)	**97.8%**	**97.1%**	**100.0%**	**98.1%**	**99.7%**	98.4%	92.7%	96.4%	97.5%
Fusion MV-HC (*top 8*)	**97.8%**	**97.1%**	**100.0%**	**98.1%**	**99.7%**	99.1%	92.3%	**96.5%**	**97.6%**
Fusion MVCP-BM (*top 9*)	95.7%	96.1%	**100.0%**	**98.1%**	99.6%	99.0%	**92.8%**	96.4%	97.2%
Fusion MVCP-HC (*top 9*)	**97.8%**	96.1%	**100.0%**	**98.1%**	99.6%	**99.3%**	92.5%	96.3%	97.5%

Results

Comparison of the recognition rates achieved across eight popular datasets by 12 models individually and through five different fusion strategies (intra-dataset experiments).

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	94.1%	**100.0%**	**97.5%**	98.1%	**97.5%**	82.6%	59.0%†	90.8%
CRNN	93.5%	88.2%	91.7%	90.7%	97.1%	92.9%	68.9%	73.6%	87.1%
Fast-OCR	93.5%	**97.1%**	**100.0%**	**97.5%**	98.1%	97.1%	81.6%	56.7%†	90.2%
GRCNN	93.5%	92.2%	93.3%	91.9%	97.1%	93.4%	66.6%	77.6%	88.2%
Holistic-CNN	87.0%	75.5%	88.3%	95.0%	97.7%	95.6%	81.2%	94.7%	89.4%
Multi-Task-LR	89.1%	73.5%	85.0%	92.5%	94.9%	93.3%	72.3%	86.6%	85.9%
R^2AM	89.1%	83.3%	86.7%	91.9%	96.5%	92.0%	75.9%	83.4%	87.4%
RARE	95.7%	94.1%	95.0%	94.4%	97.7%	94.0%	75.7%	78.7%	90.7%
Rosetta	89.1%	82.4%	93.3%	93.8%	97.5%	94.4%	75.5%	89.0%	89.4%
STAR-Net	95.7%	96.1%	95.0%	95.7%	97.8%	96.1%	78.8%	82.3%	92.2%
TRBA	93.5%	91.2%	91.7%	93.8%	97.2%	97.3%	83.4%	80.6%	91.1%
ViTSTR-Base	87.0%	88.2%	86.7%	96.9%	**99.4%**	95.8%	**89.7%**	**95.6%**	**92.4%**
Fusion HC (*top 6*)	**97.8%**	95.1%	96.7%	**98.1%**	99.0%	96.6%	90.9%	93.5%	96.0%
Fusion MV-BM (*top 8*)	**97.8%**	**97.1%**	**100.0%**	**98.1%**	**99.7%**	98.4%	92.7%	96.4%	97.5%
Fusion MV-HC (*top 8*)	**97.8%**	**97.1%**	**100.0%**	**98.1%**	**99.7%**	99.1%	92.3%	**96.5%**	**97.6%**
Fusion MVCP-BM (*top 9*)	95.7%	96.1%	**100.0%**	**98.1%**	99.6%	99.0%	**92.8%**	96.4%	97.2%
Fusion MVCP-HC (*top 9*)	**97.8%**	96.1%	**100.0%**	**98.1%**	99.6%	**99.3%**	92.5%	96.3%	97.5%

Results

While each model individually obtained recognition rates below 90% for at least two datasets, all fusion strategies surpassed the 90% threshold across all datasets.

Model \ Test set	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CR-NET	**97.8%**	94.1%	**100.0%**	**97.5%**	98.1%	**97.5%**	82.6%	59.0%†	90.8%
CRNN	93.5%	88.2%	91.7%	90.7%	97.1%	92.9%	68.9%	73.6%	87.1%
Fast-OCR	93.5%	**97.1%**	**100.0%**	**97.5%**	98.1%	97.1%	81.6%	56.7%†	90.2%
GRCNN	93.5%	92.2%	93.3%	91.9%	97.1%	93.4%	66.6%	77.6%	88.2%
Holistic-CNN	87.0%	75.5%	88.3%	95.0%	97.7%	95.6%	81.2%	94.7%	89.4%
Multi-Task-LR	89.1%	73.5%	85.0%	92.5%	94.9%	93.3%	72.3%	86.6%	85.9%
R^2AM	89.1%	83.3%	86.7%	91.9%	96.5%	92.0%	75.9%	83.4%	87.4%
RARE	95.7%	94.1%	95.0%	94.4%	97.7%	94.0%	75.7%	78.7%	90.7%
Rosetta	89.1%	82.4%	93.3%	93.8%	97.5%	94.4%	75.5%	89.0%	89.4%
STAR-Net	95.7%	96.1%	95.0%	95.7%	97.8%	96.1%	78.8%	82.3%	92.2%
TRBA	93.5%	91.2%	91.7%	93.8%	97.2%	97.3%	83.4%	80.6%	91.1%
ViTSTR-Base	87.0%	88.2%	86.7%	96.9%	**99.4%**	95.8%	**89.7%**	**95.6%**	**92.4%**
Fusion HC (*top 6*)	**97.8%**	95.1%	96.7%	**98.1%**	99.0%	96.6%	90.9%	93.5%	96.0%
Fusion MV-BM (*top 8*)	**97.8%**	**97.1%**	**100.0%**	**98.1%**	**99.7%**	98.4%	92.7%	96.4%	97.5%
Fusion MV-HC (*top 8*)	**97.8%**	**97.1%**	**100.0%**	**98.1%**	**99.7%**	99.1%	92.3%	**96.5%**	**97.6%**
Fusion MVCP-BM (*top 9*)	95.7%	96.1%	**100.0%**	**98.1%**	99.6%	99.0%	**92.8%**	96.4%	97.2%
Fusion MVCP-HC (*top 9*)	**97.8%**	96.1%	**100.0%**	**98.1%**	99.6%	**99.3%**	92.5%	96.3%	97.5%

Results (Qualitative)

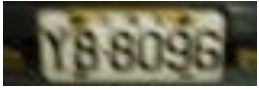
ViTSTR-Base: AIQ1Q56 (0.93)
STAR-Net: ATQ1056 (0.59)
TRBA: AIQ1056 (0.98)
CR-NET: AIQ1056 (0.82)
RARE: AIQ1Q56 (0.92)
Fusion MV-HC: AIQ1056

ViTSTR-Base: AS5I8D (0.53)
STAR-Net: AS518O (0.82)
TRBA: AS518O (0.60)
CR-NET: AS518D (0.83)
RARE: AS5I8D (0.79)
Fusion MV-HC: AS5I8D

ViTSTR-Base: 4NIU770 (0.45)
STAR-Net: 4NIU770 (0.94)
TRBA: 4NTU770 (0.99)
CR-NET: 4NTU770 (0.91)
RARE: 4NIU770 (0.99)
Fusion MV-HC: 4NIU770

ViTSTR-Base: 5EZZ29 (0.51)
STAR-Net: SEZ229 (0.74)
TRBA: 5EZ229 (0.99)
CR-NET: 5EZ229 (0.88)
RARE: 5EZ229 (0.88)
Fusion MV-HC: 5EZ229

ViTSTR-Base: KRM7E95 (0.99)
STAR-Net: KRH7E95 (0.59)
TRBA: KRM7E95 (0.51)
CR-NET: KRH7E95 (0.73)
RARE: KRM7E95 (0.60)
Fusion MV-HC: KRM7E95



ViTSTR-Base: Y88096 (0.94)
STAR-Net: Y68096 (0.93)
TRBA: Y88096 (0.97)
CR-NET: Y96096 (0.75)
RARE: YS8096 (0.67)
Fusion MV-HC: Y88096

ViTSTR-Base: HLP459A (0.98)
STAR-Net: HLP4594 (0.97)
TRBA: HLPA594 (0.99)
CR-NET: HLP4594 (0.85)
RARE: HLPA59A (0.93)
Fusion MV-HC: HLP4594

ViTSTR-Base: MRU3095 (0.97)
STAR-Net: MR03095 (0.98)
TRBA: MRD3095 (0.72)
CR-NET: MRD3095 (0.94)
RARE: MRD3095 (0.87)
Fusion MV-HC: MRD3095

Predictions obtained using multiple models individually and through the best fusion approach.

Results (Qualitative)

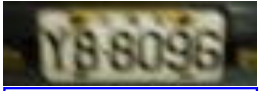
ViTSTR-Base: AIQ1Q56 (0.93)
STAR-Net: ATQ1056 (0.59)
TRBA: AIQ1056 (0.98)
CR-NET: AIQ1056 (0.82)
RARE: AIQ1Q56 (0.92)
Fusion MV-HC: AIQ1056

ViTSTR-Base: AS5I8D (0.53)
STAR-Net: AS518O (0.82)
TRBA: AS518O (0.60)
CR-NET: AS518D (0.83)
RARE: AS5I8D (0.79)
Fusion MV-HC: AS5I8D

ViTSTR-Base: 4NIU770 (0.45)
STAR-Net: 4NIU770 (0.94)
TRBA: 4NTU770 (0.99)
CR-NET: 4NTU770 (0.91)
RARE: 4NIU770 (0.99)
Fusion MV-HC: 4NIU770

ViTSTR-Base: 5EZZ29 (0.51)
STAR-Net: SEZ229 (0.74)
TRBA: 5EZ229 (0.99)
CR-NET: 5EZ229 (0.88)
RARE: 5EZ229 (0.88)
Fusion MV-HC: 5EZ229

ViTSTR-Base: KRM7E95 (0.99)
STAR-Net: KRH7E95 (0.59)
TRBA: KRM7E95 (0.51)
CR-NET: KRH7E95 (0.73)
RARE: KRM7E95 (0.60)
Fusion MV-HC: KRM7E95



ViTSTR-Base: Y88096 (0.94)
STAR-Net: Y68096 (0.93)
TRBA: Y88096 (0.97)
CR-NET: Y96096 (0.75)
RARE: YS8096 (0.67)
Fusion MV-HC: Y88096

ViTSTR-Base: HLP459A (0.98)
STAR-Net: HLP4594 (0.97)
TRBA: HLPA594 (0.99)
CR-NET: HLP4594 (0.85)
RARE: HLPA59A (0.93)
Fusion MV-HC: HLP4594

ViTSTR-Base: MRU3095 (0.97)
STAR-Net: MR03095 (0.98)
TRBA: MRD3095 (0.72)
CR-NET: MRD3095 (0.94)
RARE: MRD3095 (0.87)
Fusion MV-HC: MRD3095

Model fusion can produce accurate predictions even in cases where most models exhibit prediction errors.

Results

Average results obtained across the datasets by combining the output of the top N models.

OCR Models	HC	MV-BM	MV-HC	MVCP-BM	MVCP-HC
Top 1 (ViTSTR-Base)	92.4%	92.4%	92.4%	92.4%	92.4%
Top 2 (+ STAR-Net)	94.1%	92.4%	94.1%	92.4%	94.1%
Top 3 (+ TRBA)	94.2%	94.6%	94.9%	94.2%	94.2%
Top 4 (+ CR-NET)	95.2%	95.9%	96.3%	94.8%	95.9%
Top 5 (+ RARE)	95.5%	96.1%	96.6%	96.1%	96.2%
Top 6 (+ Fast-OCR)	**96.0%**	97.1%	97.0%	96.7%	96.9%
Top 7 (+ Rosetta)	95.4%	97.3%	97.2%	97.1%	97.0%
Top 8 (+ Holistic-CNN)	95.7%	**97.5%**	**97.6%**	96.1%	97.2%
Top 9 (+ GRCNN)	95.7%	97.5%	97.5%	**97.2%**	**97.5%**
Top 10 (+ R^2AM)	95.5%	97.4%	97.2%	96.1%	96.6%
Top 11 (+ CRNN)	95.2%	97.1%	97.0%	96.5%	96.5%
Top 12 (+ Multi-Task-LR)	95.0%	97.0%	97.0%	95.5%	96.5%

- The best results were reached using the sequence-level majority vote approaches (MV-*).

Results

Average results obtained across the datasets by combining the output of the top N models.

OCR Models	HC	MV-BM	MV-HC	MVCP-BM	MVCP-HC
Top 1 (ViTSTR-Base)	92.4%	92.4%	92.4%	92.4%	92.4%
Top 2 (+ STAR-Net)	94.1%	92.4%	94.1%	92.4%	94.1%
Top 3 (+ TRBA)	94.2%	94.6%	94.9%	94.2%	94.2%
Top 4 (+ CR-NET)	95.2%	95.9%	96.3%	94.8%	95.9%
Top 5 (+ RARE)	95.5%	96.1%	96.6%	96.1%	96.2%
Top 6 (+ Fast-OCR)	**96.0%**	97.1%	97.0%	96.7%	96.9%
Top 7 (+ Rosetta)	95.4%	97.3%	97.2%	97.1%	97.0%
Top 8 (+ Holistic-CNN)	95.7%	**97.5%**	**97.6%**	96.1%	97.2%
Top 9 (+ GRCNN)	95.7%	97.5%	97.5%	**97.2%**	**97.5%**
Top 10 (+ R^2AM)	95.5%	97.4%	97.2%	96.1%	96.6%
Top 11 (+ CRNN)	95.2%	97.1%	97.0%	96.5%	96.5%
Top 12 (+ Multi-Task-LR)	95.0%	97.0%	97.0%	95.5%	96.5%

- Selecting the prediction with the highest confidence (HC) led to suboptimal results.
  - All models tend to make incorrect predictions also with high confidence.

Results (Cross-Dataset)

And in **cross-dataset** scenarios?

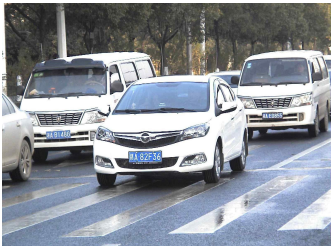
Results (Cross-Dataset)

And in **cross-dataset** scenarios?

OpenALPR-EU + 3 new datasets:

PKU (2017)

CD-HARD (2018)



CLPD (2021)

Results (Cross-Dataset)

Results achieved in cross-dataset setups.

Approach \ Test Dataset	OpenALPR-EU # 108	PKU # 2,253	CD-HARD # 104	CLPD # 1,200	Average
CR-NET	96.3%	99.1%	58.7%	94.2%	87.1%
CRNN	93.5%	98.2%	31.7%	89.0%	78.1%
Fast-OCR	**97.2%**	**99.2%**	**59.6%**	**94.4%**	**87.6%**
GRCNN	87.0%	98.6%	38.5%	87.7%	77.9%
Holistic-CNN	89.8%	98.6%	11.5%	90.2%	72.5%
Multi-Task-LR	85.2%	97.4%	10.6%	86.8%	70.0%
R^2AM	88.9%	97.1%	20.2%	88.2%	73.6%
RARE	94.4%	98.3%	37.5%	92.4%	80.7%
Rosetta	90.7%	97.2%	14.4%	86.9%	72.3%
STAR-Net	**97.2%**	99.1%	48.1%	93.3%	84.4%
TRBA	93.5%	98.5%	35.6%	90.9%	79.6%
ViTSTR-Base	89.8%	98.4%	22.1%	93.1%	75.9%
Fusion HC (*top 6*)	95.4%	99.2%	48.1%	94.9%	84.4%
Fusion MV–BM (*top 8*)	**99.1%**	**99.7%**	**65.4%**	**97.0%**	**90.3%**
Fusion MV–HC (*top 8*)	**99.1%**	**99.7%**	**65.4%**	96.3%	90.1%
Fusion MVCP–BM (*top 9*)	95.4%	**99.7%**	54.8%	95.5%	86.3%
Fusion MVCP–HC (*top 9*)	97.2%	**99.7%**	57.7%	95.9%	87.6%

Results (Cross-Dataset)

Results achieved in cross-dataset setups.

Approach \ Test Dataset	OpenALPR-EU # 108	PKU # 2,253	CD-HARD # 104	CLPD # 1,200	Average
CR-NET	96.3%	99.1%	58.7%	94.2%	87.1%
CRNN	93.5%	98.2%	31.7%	89.0%	78.1%
Fast-OCR	**97.2%**	**99.2%**	**59.6%**	**94.4%**	**87.6%**
GRCNN	87.0%	98.6%	38.5%	87.7%	77.9%
Holistic-CNN	89.8%	98.6%	11.5%	90.2%	72.5%
Multi-Task-LR	85.2%	97.4%	10.6%	86.8%	70.0%
R^2AM	88.9%	97.1%	20.2%	88.2%	73.6%
RARE	94.4%	98.3%	37.5%	92.4%	80.7%
Rosetta	90.7%	97.2%	14.4%	86.9%	72.3%
STAR-Net	**97.2%**	99.1%	48.1%	93.3%	84.4%
TRBA	93.5%	98.5%	35.6%	90.9%	79.6%
ViTSTR-Base	89.8%	98.4%	22.1%	93.1%	75.9%
Fusion HC (*top 6*)	95.4%	99.2%	48.1%	94.9%	84.4%
Fusion MV–BM (*top 8*)	**99.1%**	**99.7%**	**65.4%**	**97.0%**	**90.3%**
Fusion MV–HC (*top 8*)	**99.1%**	**99.7%**	**65.4%**	96.3%	90.1%
Fusion MVCP–BM (*top 9*)	95.4%	**99.7%**	54.8%	95.5%	86.3%
Fusion MVCP–HC (*top 9*)	97.2%	**99.7%**	57.7%	95.9%	87.6%

Results (Cross-Dataset)

Results achieved in cross-dataset setups.

Approach \ Test Dataset	OpenALPR-EU # 108	PKU # 2,253	CD-HARD # 104	CLPD # 1,200	Average
CR-NET	96.3%	99.1%	58.7%	94.2%	87.1%
CRNN	93.5%	98.2%	31.7%	89.0%	78.1%
Fast-OCR	**97.2%**	**99.2%**	**59.6%**	**94.4%**	**87.6%**
GRCNN	87.0%	98.6%	38.5%	87.7%	77.9%
Holistic-CNN	89.8%	98.6%	11.5%	90.2%	72.5%
Multi-Task-LR	85.2%	97.4%	10.6%	86.8%	70.0%
R^2AM	88.9%	97.1%	20.2%	88.2%	73.6%
RARE	94.4%	98.3%	37.5%	92.4%	80.7%
Rosetta	90.7%	97.2%	14.4%	86.9%	72.3%
STAR-Net	**97.2%**	99.1%	48.1%	93.3%	84.4%
TRBA	93.5%	98.5%	35.6%	90.9%	79.6%
ViTSTR-Base	89.8%	98.4%	22.1%	93.1%	75.9%
Fusion HC (*top 6*)	95.4%	99.2%	48.1%	94.9%	84.4%
Fusion MV–BM (*top 8*)	**99.1%**	**99.7%**	**65.4%**	**97.0%**	**90.3%**
Fusion MV–HC (*top 8*)	**99.1%**	**99.7%**	**65.4%**	96.3%	90.1%
Fusion MVCP–BM (*top 9*)	95.4%	**99.7%**	54.8%	95.5%	86.3%
Fusion MVCP–HC (*top 9*)	97.2%	**99.7%**	57.7%	95.9%	87.6%

Results (Speed/Accuracy Trade-Off)

The number of FPS processed by each model independently and when incorporated into the ensembles. The reported time, measured in milliseconds per image, represents the average of 5 runs.

Models (ranked by **accuracy**)	MV–HC	Individual		Fusion	
		Time	FPS	Time	FPS
Top 1 (ViTSTR-Base)	92.4%	7.3	137	7.3	137
Top 2 (+ STAR-Net)	94.1%	7.1	141	14.4	70
Top 3 (+ TRBA)	94.9%	16.9	59	31.3	32
Top 4 (+ CR-NET)	96.3%	5.3	189	36.6	27
Top 5 (+ RARE)	96.6%	13.0	77	49.6	20
Top 6 (+ Fast-OCR)	97.0%	3.0	330	52.6	19
Top 7 (+ Rosetta)	97.2%	4.6	219	57.2	18
Top 8 (+ Holistic-CNN)	97.6%	2.5	399	59.7	17
Top 9 (+ GRCNN)	97.5%	8.5	117	68.2	15
Top 10 (+ R^2AM)	97.2%	15.9	63	84.2	12
Top 11 (+ CRNN)	97.0%	2.9	343	87.1	11
Top 12 (+ Multi-Task-LR)	97.0%	2.3	427	89.4	11

Models (ranked by **speed**)	MV–HC	Individual		Fusion	
		Time	FPS	Time	FPS
Top 1 (Multi-Task-LR)	85.9%	2.3	427	2.3	427
Top 2 (+ Holistic-CNN)	90.2%	2.5	399	4.9	206
Top 3 (+ CRNN)	91.1%	2.9	343	7.8	129
Top 4 (+ Fast-OCR)	95.4%	3.0	330	10.8	93
Top 5 (+ Rosetta)	96.0%	4.6	219	15.4	65
Top 6 (+ CR-NET)	96.6%	5.3	189	20.7	48
Top 7 (+ STAR-Net)	96.9%	7.1	141	27.8	36
Top 8 (+ ViTSTR-Base)	96.9%	7.3	137	35.0	29
Top 9 (+ GRCNN)	97.1%	8.5	117	43.6	23
Top 10 (+ RARE)	97.1%	13.0	77	56.6	18
Top 11 (+ R^2AM)	97.1%	15.9	63	72.5	14
Top 12 (+ TRBA)	97.1%	16.9	59	89.4	11

- All experiments were conducted using an *NVIDIA Quadro RTX* 8000 GPU.

Results (Speed/Accuracy Trade-Off)

The number of FPS processed by each model independently and when incorporated into the ensembles. The reported time, measured in milliseconds per image, represents the average of 5 runs.

Models (ranked by **accuracy**)	MV–HC	Individual		Fusion	
		Time	FPS	Time	FPS
Top 1 (ViTSTR-Base)	92.4%	7.3	137	7.3	137
Top 2 (+ STAR-Net)	94.1%	7.1	141	14.4	70
Top 3 (+ TRBA)	94.9%	16.9	59	31.3	32
Top 4 (+ CR-NET)	96.3%	5.3	189	36.6	27
Top 5 (+ RARE)	96.6%	13.0	77	49.6	20
Top 6 (+ Fast-OCR)	97.0%	3.0	330	52.6	19
Top 7 (+ Rosetta)	97.2%	4.6	219	57.2	18
Top 8 (+ Holistic-CNN)	97.6%	2.5	399	59.7	17
Top 9 (+ GRCNN)	97.5%	8.5	117	68.2	15
Top 10 (+ R^2AM)	97.2%	15.9	63	84.2	12
Top 11 (+ CRNN)	97.0%	2.9	343	87.1	11
Top 12 (+ Multi-Task-LR)	97.0%	2.3	427	89.4	11

Models (ranked by **speed**)	MV–HC	Individual		Fusion	
		Time	FPS	Time	FPS
Top 1 (Multi-Task-LR)	85.9%	2.3	427	2.3	427
Top 2 (+ Holistic-CNN)	90.2%	2.5	399	4.9	206
Top 3 (+ CRNN)	91.1%	2.9	343	7.8	129
Top 4 (+ Fast-OCR)	95.4%	3.0	330	10.8	93
Top 5 (+ Rosetta)	96.0%	4.6	219	15.4	65
Top 6 (+ CR-NET)	96.6%	5.3	189	20.7	48
Top 7 (+ STAR-Net)	96.9%	7.1	141	27.8	36
Top 8 (+ ViTSTR-Base)	96.9%	7.3	137	35.0	29
Top 9 (+ GRCNN)	97.1%	8.5	117	43.6	23
Top 10 (+ RARE)	97.1%	13.0	77	56.6	18
Top 11 (+ R^2AM)	97.1%	15.9	63	72.5	14
Top 12 (+ TRBA)	97.1%	16.9	59	89.4	11

- Fusing the outputs of the three fastest models results in a lower recognition rate (91.1%) than using the best model alone (92.4%).

Results (Speed/Accuracy Trade-Off)

The number of FPS processed by each model independently and when incorporated into the ensembles. The reported time, measured in milliseconds per image, represents the average of 5 runs.

Models (ranked by **accuracy**)	MV–HC	Individual		Fusion	
		Time	FPS	Time	FPS
Top 1 (ViTSTR-Base)	92.4%	7.3	137	7.3	137
Top 2 (+ STAR-Net)	94.1%	7.1	141	14.4	70
Top 3 (+ TRBA)	94.9%	16.9	59	31.3	32
Top 4 (+ CR-NET)	96.3%	5.3	189	36.6	27
Top 5 (+ RARE)	96.6%	13.0	77	49.6	20
Top 6 (+ Fast-OCR)	97.0%	3.0	330	52.6	19
Top 7 (+ Rosetta)	97.2%	4.6	219	57.2	18
Top 8 (+ Holistic-CNN)	97.6%	2.5	399	59.7	17
Top 9 (+ GRCNN)	97.5%	8.5	117	68.2	15
Top 10 (+ R^2AM)	97.2%	15.9	63	84.2	12
Top 11 (+ CRNN)	97.0%	2.9	343	87.1	11
Top 12 (+ Multi-Task-LR)	97.0%	2.3	427	89.4	11

Models (ranked by **speed**)	MV–HC	Individual		Fusion	
		Time	FPS	Time	FPS
Top 1 (Multi-Task-LR)	85.9%	2.3	427	2.3	427
Top 2 (+ Holistic-CNN)	90.2%	2.5	399	4.9	206
Top 3 (+ CRNN)	91.1%	2.9	343	7.8	129
Top 4 (+ Fast-OCR)	95.4%	3.0	330	10.8	93
Top 5 (+ Rosetta)	96.0%	4.6	219	15.4	65
Top 6 (+ CR-NET)	96.6%	5.3	189	20.7	48
Top 7 (+ STAR-Net)	96.9%	7.1	141	27.8	36
Top 8 (+ ViTSTR-Base)	96.9%	7.3	137	35.0	29
Top 9 (+ GRCNN)	97.1%	8.5	117	43.6	23
Top 10 (+ RARE)	97.1%	13.0	77	56.6	18
Top 11 (+ R^2AM)	97.1%	15.9	63	72.5	14
Top 12 (+ TRBA)	97.1%	16.9	59	89.4	11

- Combining **4-6 fast models** appears to be the optimal choice for striking a better balance between speed and accuracy.

Highlights

- **Substantial benefits of fusion approaches** in both intra- and cross-dataset setups;
  - Optimal fusion approach $\rightarrow$ Majority Vote at the sequence level;
  - Intra-dataset: 92.4% $\rightarrow$ 97.6% || Cross-dataset: 87.6% $\rightarrow$ 90.3%;

Highlights

- **Substantial benefits of fusion approaches** in both intra- and cross-dataset setups;
  - Optimal fusion approach $\rightarrow$ Majority Vote at the sequence level;
  - Intra-dataset: 92.4% $\rightarrow$ 97.6% || Cross-dataset: 87.6% $\rightarrow$ 90.3%;
- For applications where the recognition task can tolerate some additional time, though not excessively, **an effective strategy is to combine 4-6 fast models**.
  - These 4-6 models may not be the most accurate individually, but their fusion strikes an appealing balance between speed and accuracy.

Recap – Hypothesis and Research Questions

Hypothesis

It is possible to significantly improve the state of the art in ALPR without increasing the number of real training images, designing groundbreaking descriptors, or extensively searching for better model architectures.

Some questions that guide our research are:

- How can we address the lack of attention given to images featuring Mercosur LPs?
- Do current methods for detecting and recognizing LPs generalize well to unseen data?
- Can we significantly improve results by combining the outputs of various OCR models?
- To what extent does combining real data with synthetic data improve LPR accuracy?

Advancing Multinational License Plate Recognition Through Synthetic and Real Data Fusion: A Comprehensive Evaluation

Synthetic Data – Templates

Examples of template-based LP images we created for this study.

Synthetic Data – Character Permutation

Examples of LP images created via character permutation.

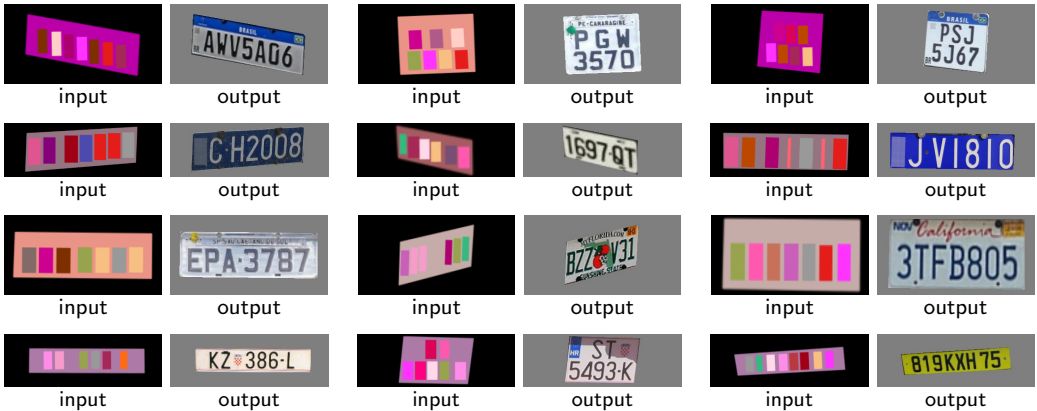
Synthetic Data – Generative Adversarial Network (GAN)

pix2pix (https://phillipi.github.io/pix2pix/)

- Paired image-to-image translation:

Synthetic Data – GAN Training

Examples of image pairs used for **training** the pix2pix model:



Synthetic Data – GAN

Not all images meet satisfactory quality standards!

Examples of well-generated images:

Examples of poorly generated images:

Synthetic Data – GAN

Not all images meet satisfactory quality standards!

Examples of well-generated images:

Examples of poorly generated images:

We applied the Fast-OCR model to distinguish between well and poorly generated images.

Synthetic Data – GAN

Examples of **selected images** from those generated using pix2pix:

Results

All 16 models exhibited **exceptional performance**!

Model \ Test set # LPs	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CNNG (Fan and Zhao, 2022)	**97.8%**	91.2%	96.7%	98.8%	99.1%	98.8%	**96.1%**	97.1%	96.9%
CR-NET (Silva and Jung, 2020)	93.5%	96.1%	98.3%	96.9%	98.7%	98.0%	89.3%	88.3%†	94.9%
CRNN (Shi et al., 2017)	93.5%	96.1%	96.7%	95.7%	98.8%	97.5%	87.0%	92.2%	94.7%
Fast-OCR (Laroca et al., 2021a)	95.7%	97.1%	95.0%	96.9%	98.7%	96.0%	89.6%	88.1%†	94.6%
GRCNN (Wang and Hu, 2017)	**97.8%**	**99.0%**	96.7%	98.8%	99.0%	97.9%	87.4%	93.0%	96.2%
Holistic-CNN (Špaňhel et al., 2017)	95.7%	91.2%	93.3%	99.4%	99.3%	98.4%	94.9%	**97.9%**	96.3%
Multi-Task (Gonçalves et al., 2018)	**97.8%**	94.1%	**100.0%**	98.8%	99.1%	98.6%	93.3%	95.1%	97.1%
Multi-Task-LR (Gonçalves et al., 2019)	95.7%	93.1%	93.3%	**100.0%**	99.6%	97.5%	94.6%	96.6%	96.3%
R^2AM (Lee and Osindero, 2016)	**97.8%**	94.1%	95.0%	98.8%	99.3%	99.3%	90.6%	94.4%	96.1%
RARE (Shi et al., 2016)	**97.8%**	97.1%	98.3%	98.1%	99.4%	99.1%	91.9%	96.5%	97.3%
Rosetta (Borisyuk et al., 2018)	95.7%	98.0%	98.3%	98.1%	98.7%	98.3%	92.6%	96.0%	97.0%
STAR-Net (Liu et al., 2016b)	**97.8%**	**99.0%**	98.3%	98.1%	99.1%	99.3%	94.7%	97.0%	**97.9%**
TRBA (Baek et al., 2019)	**97.8%**	**99.0%**	98.3%	98.8%	98.8%	99.3%	94.0%	97.3%	**97.9%**
ViTSTR-Base (Atienza, 2021b)	95.7%	96.1%	93.3%	99.4%	**99.9%**	**99.4%**	94.6%	97.7%	97.0%
ViTSTR-Small (Atienza, 2021b)	95.7%	96.1%	98.3%	98.1%	99.1%	98.5%	94.9%	96.8%	97.2%
ViTSTR-Tiny (Atienza, 2021b)	93.5%	95.1%	91.7%	98.8%	99.0%	98.9%	92.3%	95.3%	95.5%
Average	96.2%	95.8%	96.4%	98.3%	99.1%	98.4%	92.4%	94.9%	96.4%

†Images from the RodoSol-ALPR dataset were not used for training the CR-NET and Fast-OCR models, as each character's bounding box needs to be labeled for training them.

Results

All 16 models exhibited **exceptional performance**!

Model \ Test set # LPs	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CNNG (Fan and Zhao, 2022)	**97.8%**	91.2%	96.7%	98.8%	99.1%	98.8%	**96.1%**	97.1%	96.9%
CR-NET (Silva and Jung, 2020)	93.5%	96.1%	98.3%	96.9%	98.7%	98.0%	89.3%	88.3%†	94.9%
CRNN (Shi et al., 2017)	93.5%	96.1%	96.7%	95.7%	98.8%	97.5%	87.0%	92.2%	94.7%
Fast-OCR (Laroca et al., 2021a)	95.7%	97.1%	95.0%	96.9%	98.7%	96.0%	89.6%	88.1%†	94.6%
GRCNN (Wang and Hu, 2017)	**97.8%**	**99.0%**	96.7%	98.8%	99.0%	97.9%	87.4%	93.0%	96.2%
Holistic-CNN (Špaňhel et al., 2017)	95.7%	91.2%	93.3%	99.4%	99.3%	98.4%	94.9%	**97.9%**	96.3%
Multi-Task (Gonçalves et al., 2018)	**97.8%**	94.1%	**100.0%**	98.8%	99.1%	98.6%	93.3%	95.1%	97.1%
Multi-Task-LR (Gonçalves et al., 2019)	95.7%	93.1%	93.3%	**100.0%**	99.6%	97.5%	94.6%	96.6%	96.3%
R^2AM (Lee and Osindero, 2016)	**97.8%**	94.1%	95.0%	98.8%	99.3%	99.3%	90.6%	94.4%	96.1%
RARE (Shi et al., 2016)	**97.8%**	97.1%	98.3%	98.1%	99.4%	99.1%	91.9%	96.5%	97.3%
Rosetta (Borisyuk et al., 2018)	95.7%	98.0%	98.3%	98.1%	98.7%	98.3%	92.6%	96.0%	97.0%
STAR-Net (Liu et al., 2016b)	**97.8%**	**99.0%**	98.3%	98.1%	99.1%	99.3%	94.7%	97.0%	**97.9%**
TRBA (Baek et al., 2019)	**97.8%**	**99.0%**	98.3%	98.8%	98.8%	99.3%	94.0%	97.3%	**97.9%**
ViTSTR-Base (Atienza, 2021b)	95.7%	96.1%	93.3%	99.4%	**99.9%**	**99.4%**	94.6%	97.7%	97.0%
ViTSTR-Small (Atienza, 2021b)	95.7%	96.1%	98.3%	98.1%	99.1%	98.5%	94.9%	96.8%	97.2%
ViTSTR-Tiny (Atienza, 2021b)	93.5%	95.1%	91.7%	98.8%	99.0%	98.9%	92.3%	95.3%	95.5%
Average	96.2%	95.8%	96.4%	98.3%	99.1%	98.4%	92.4%	94.9%	96.4%

†Images from the RodoSol-ALPR dataset were not used for training the CR-NET and Fast-OCR models, as each character's bounding box needs to be labeled for training them.

Results

All 16 models exhibited **exceptional performance**!

Model \ Test set # LPs	Caltech Cars # 46	EnglishLP # 102	UCSD-Stills # 60	ChineseLP # 161	AOLP # 687	SSIG-SegPlate # 804	UFPR-ALPR # 1,800	RodoSol-ALPR # 8,000	Average
CNNG (Fan and Zhao, 2022)	**97.8%**	91.2%	96.7%	98.8%	99.1%	98.8%	**96.1%**	97.1%	96.9%
CR-NET (Silva and Jung, 2020)	93.5%	96.1%	98.3%	96.9%	98.7%	98.0%	89.3%	88.3%†	94.9%
CRNN (Shi et al., 2017)	93.5%	96.1%	96.7%	95.7%	98.8%	97.5%	87.0%	92.2%	94.7%
Fast-OCR (Laroca et al., 2021a)	95.7%	97.1%	95.0%	96.9%	98.7%	96.0%	89.6%	88.1%†	94.6%
GRCNN (Wang and Hu, 2017)	**97.8%**	**99.0%**	96.7%	98.8%	99.0%	97.9%	87.4%	93.0%	96.2%
Holistic-CNN (Špaňhel et al., 2017)	95.7%	91.2%	93.3%	99.4%	99.3%	98.4%	94.9%	**97.9%**	96.3%
Multi-Task (Gonçalves et al., 2018)	**97.8%**	94.1%	**100.0%**	98.8%	99.1%	98.6%	93.3%	95.1%	97.1%
Multi-Task-LR (Gonçalves et al., 2019)	95.7%	93.1%	93.3%	**100.0%**	99.6%	97.5%	94.6%	96.6%	96.3%
R^2AM (Lee and Osindero, 2016)	**97.8%**	94.1%	95.0%	98.8%	99.3%	99.3%	90.6%	94.4%	96.1%
RARE (Shi et al., 2016)	**97.8%**	97.1%	98.3%	98.1%	99.4%	99.1%	91.9%	96.5%	97.3%
Rosetta (Borisyuk et al., 2018)	95.7%	98.0%	98.3%	98.1%	98.7%	98.3%	92.6%	96.0%	97.0%
STAR-Net (Liu et al., 2016b)	**97.8%**	**99.0%**	98.3%	98.1%	99.1%	99.3%	94.7%	97.0%	**97.9%**
TRBA (Baek et al., 2019)	**97.8%**	**99.0%**	98.3%	98.8%	98.8%	99.3%	94.0%	97.3%	**97.9%**
ViTSTR-Base (Atienza, 2021b)	95.7%	96.1%	93.3%	99.4%	**99.9%**	**99.4%**	94.6%	97.7%	97.0%
ViTSTR-Small (Atienza, 2021b)	95.7%	96.1%	98.3%	98.1%	99.1%	98.5%	94.9%	96.8%	97.2%
ViTSTR-Tiny (Atienza, 2021b)	93.5%	95.1%	91.7%	98.8%	99.0%	98.9%	92.3%	95.3%	95.5%
Model Fusion MV-HC (top 8)	97.8%	99.0%	100.0%	99.4%	99.6%	100.0%	98.4%	98.7%	**99.1%**

†Images from the RodoSol-ALPR dataset were not used for training the CR-NET and Fast-OCR models, as each character's bounding box needs to be labeled for training them.

Results – Synergistic Effect

Average recognition rates obtained across all models and datasets with different types of images included in the training set.

Real Images + data aug.	Templates	Permutation	GAN (pix2pix)	Average	Average (rect.)
	✓			42.5%	46.5%
✓				84.5%	88.1%
✓		✓		91.4%	93.6%
✓	✓			92.5%	94.7%
✓			✓	93.2%	95.2%
✓	✓	✓		93.8%	95.5%
✓		✓	✓	94.0%	95.6%
✓	✓		✓	94.1%	95.8%
✓	✓	✓	✓	**94.9%**	**96.4%**

Results – Synergistic Effect

Average recognition rates obtained across all models and datasets with different types of images included in the training set.

Real Images + data aug.	Templates	Permutation	GAN (pix2pix)	Average	Average (rect.)
	✓			42.5%	46.5%
✓				84.5%	88.1%
✓		✓		91.4%	93.6%
✓	✓			92.5%	94.7%
✓			✓	93.2%	95.2%
✓	✓	✓		93.8%	95.5%
✓		✓	✓	94.0%	95.6%
✓	✓		✓	94.1%	95.8%
✓	✓	✓	✓	**94.9%**	**96.4%**

Results – Limited Training Data

Average recognition rates obtained by STAR-Net and TRBA when trained with reduced portions of the original training data.

Model \ Real Images	100%	50%	25%	10%	5%	1%
STAR-Net (no synthetic)	95.3%	62.0%	18.3%	1.3%	0.2%	0.0%
STAR-Net (w/ synthetic)	97.9%	95.8%	94.7%	94.6%	93.6%	86.4%
TRBA (no synthetic)	93.7%	74.0%	23.9%	0.9%	0.2%	0.0%
TRBA (w/ synthetic)	97.9%	97.0%	96.0%	94.5%	94.3%	87.9%

Results – Limited Training Data

Average recognition rates obtained by STAR-Net and TRBA when trained with reduced portions of the original training data.

Model \ Real Images	100%	50%	25%	10%	5%	1%
STAR-Net (no synthetic)	95.3%	62.0%	18.3%	1.3%	0.2%	0.0%
STAR-Net (w/ synthetic)	97.9%	95.8%	94.7%	94.6%	93.6%	86.4%
TRBA (no synthetic)	93.7%	74.0%	23.9%	0.9%	0.2%	0.0%
TRBA (w/ synthetic)	97.9%	97.0%	96.0%	94.5%	94.3%	87.9%

Results (Cross-Dataset)

Comparison of the recognition rates obtained by our best approach, state-of-the-art methods, and commercial systems on the CLPD and PKU datasets (cross-dataset experiments).

Approach	Real images of Chinese LPs used for training	Multinational	Recognition Rate CLPD	Recognition Rate PKU
Sighthound (2023)	?	✓	85.2%	89.3%
Zhang et al. (2021c)	100,000+		87.6%	90.5%
Fan and Zhao (2022)	100,000+	✓	88.5%	92.5%
Ours	506	✓	90.1%	96.8%
Rao et al. (2024)	4,444		91.4%	96.1%
Liu et al. (2021)	10,000		91.7%	–
OpenALPR (2023)	?		91.8%	96.0%
Chen et al. (2023)	100,000+		92.4%	92.8%
Ke et al. (2023)	100,000+		93.2%	–
Zou et al. (2020)	100,000+		94.0%	96.6%
Zou et al. (2022)	100,000+		94.5%	–
Wang et al. (2022b)	100,000+		94.8%	–
Wang et al. (2022c)	100,000+		95.3%	96.9%
Ours + synthetic	**506**	✓	**96.2%**	**99.4%**

Results (Cross-Dataset)

Comparison of the recognition rates obtained by our best approach, state-of-the-art methods, and commercial systems on the CLPD and PKU datasets (cross-dataset experiments).

Approach	Real images of Chinese LPs used for training	Multinational	Recognition Rate CLPD	Recognition Rate PKU
Sighthound (2023)	?	✓	85.2%	89.3%
Zhang et al. (2021c)	100,000+		87.6%	90.5%
Fan and Zhao (2022)	100,000+	✓	88.5%	92.5%
Ours	506	✓	90.1%	96.8%
Rao et al. (2024)	4,444		91.4%	96.1%
Liu et al. (2021)	10,000		91.7%	–
OpenALPR (2023)	?		91.8%	96.0%
Chen et al. (2023)	100,000+		92.4%	92.8%
Ke et al. (2023)	100,000+		93.2%	–
Zou et al. (2020)	100,000+		94.0%	96.6%
Zou et al. (2022)	100,000+		94.5%	–
Wang et al. (2022b)	100,000+		94.8%	–
Wang et al. (2022c)	100,000+		95.3%	96.9%
Ours + synthetic	**506**	✓	**96.2%**	**99.4%**
Ours + synthetic + model fusion	**506**	✓	**97.6%**	**99.6%**

Highlights

- A **synergistic effect** was observed when combining different synthesis methods;
  - State-of-the-art results in both intra- and cross-dataset scenarios;

Highlights

- A **synergistic effect** was observed when combining different synthesis methods;
  - State-of-the-art results in both intra- and cross-dataset scenarios;
- Synthetic LP images proved **highly effective** in overcoming the challenges posed by limited training data;
  - Commendable results were attained even when using small fractions of the original data.

Highlights

- A **synergistic effect** was observed when combining different synthesis methods;
  - State-of-the-art results in both intra- and cross-dataset scenarios;
- Synthetic LP images proved **highly effective** in overcoming the challenges posed by limited training data;
  - Commendable results were attained even when using small fractions of the original data.
- **Even better results** are achieved by exploring both synthetic data and model fusion.

Automatic License Plate Recognition (ALPR): Toward Improving the State of the Art and Bridging the Gap Between Academia and Industry

Do We Train on Test Data?
The Impact of Near-Duplicates on License Plate Recognition

Near-Duplicates – AOLP dataset

(a) Subset AC

(b) Subset LE

(c) Subset RP

(d) Subset AC

(e) Subset AC

(f) Subset RP

In the split protocols traditionally adopted in the literature,
some of these images are in the training set and others are in the test set.

Near-Duplicates – CCPD dataset


Subset Base


Subset Base


Subset Base


Subset Base

(a) Training set


Subset Challenge


Subset Challenge


Subset Weather


Subset Weather

(b) Test set

Research Question

Research Question

To what extent have such near-duplicates impacted the evaluation of deep learning-based models applied to LPR?

Experimental Setup

We explored the two most popular datasets in the field:

- AOLP (https://github.com/avlab-cv/aolp);
- CCPD (https://github.com/detectrecog/ccpd).

Experimental Setup

We explored the two most popular datasets in the field:

- AOLP (https://github.com/avlab-cv/aolp);
- CCPD (https://github.com/detectrecog/ccpd).

We created ***fair splits*** for each dataset, where:

- There are no duplicates in the training and test sets;
- The key characteristics of the original partitions are preserved as much as possible.

Experimental Setup

We explored the two most popular datasets in the field:

- AOLP (https://github.com/avlab-cv/aolp);
- CCPD (https://github.com/detectrecog/ccpd).

We created ***fair splits*** for each dataset, where:

- There are no duplicates in the training and test sets;
- The key characteristics of the original partitions are preserved as much as possible.

We compared the performance of **six OCR models** under the traditional (adopted in previous works) and fair protocols:

OCR Model	Original Application
CNNG	License Plate Recognition
Holistic-CNN	License Plate Recognition
Multi-Task	License Plate Recognition

OCR Model	Original Application
STAR-Net	Scene Text Recognition
TRBA	Scene Text Recognition
ViTSTR-Base	Scene Text Recognition

Results – AOLP

Results under the AOLP[3] (adopted in previous works) and AOLP-Fair (ours) protocols.

Model	AOLP ↑	AOLP-Fair ↑	Gap ↓	Rel. Gap ↓
CNNG	**98.91%**	96.80%	2.11%	193.6%
Holistic-CNN	98.42%	96.30%	2.12%	134.2%
Multi-Task	98.42%	95.29%	3.13%	198.1%
STAR-Net	98.47%	96.46%	2.01%	131.4%
TRBA	98.75%	**97.47%**	**1.28%**	**102.4%**
ViTSTR-Base	98.75%	97.31%	1.44%	115.2%

The error rates were **more than twice as high** in the experiments conducted under the fair protocol, which has no duplicates.

[3] 67.6% of the test images have duplicates in the training set.

Results – AOLP

Results under the AOLP[3] (adopted in previous works) and AOLP-Fair (ours) protocols.

Model	AOLP ↑	AOLP-Fair ↑	Gap ↓	Rel. Gap ↓
CNNG	**98.91%**	96.80%	2.11%	193.6%
Holistic-CNN	98.42%	96.30%	2.12%	134.2%
Multi-Task	98.42%	95.29%	3.13%	198.1%
STAR-Net	98.47%	96.46%	2.01%	131.4%
TRBA	98.75%	**97.47%**	**1.28%**	**102.4%**
ViTSTR-Base	98.75%	97.31%	1.44%	115.2%

The error rates were more than twice as high in the experiments conducted under the fair protocol, which has no duplicates.

The ranking of the models **changed** when they were trained and tested under fair splits.
Best model: **CNNG** → **TRBA**

[3] 67.6% of the test images have duplicates in the training set.

Results – CCPD

Results achieved on the CCPD dataset under the standard[4] and CCPD-Fair protocols.

Model	CCPD ↑	CCPD-Fair ↑	Gap ↓	Rel. Gap ↓
CNNG	**88.24%**	**86.93%**	1.31%	11.1%
Holistic-CNN	77.01%	75.41%	1.60%	7.0%
Multi-Task	83.01%	81.84%	**1.17%**	**6.9%**
STAR-Net	78.53%	73.33%	5.20%	24.2%
TRBA	75.83%	71.48%	4.35%	18.0%
ViTSTR-Base	79.06%	76.37%	2.69%	12.9%

[4]CCPD's standard protocol: 19.1% of the test images have duplicates in the training set.

Results – CCPD

Results achieved on the CCPD dataset under the standard[4] and CCPD-Fair protocols.

Model	CCPD ↑	CCPD-Fair ↑	Gap ↓	Rel. Gap ↓
CNNG	**88.24%**	**86.93%**	1.31%	11.1%
Holistic-CNN	77.01%	75.41%	1.60%	7.0%
Multi-Task	83.01%	81.84%	**1.17%**	**6.9%**
STAR-Net	78.53%	73.33%	5.20%	24.2%
TRBA	75.83%	71.48%	4.35%	18.0%
ViTSTR-Base	79.06%	76.37%	2.69%	12.9%

The CCPD dataset has $\approx$ 157K test images:

- The lowest performance gap of **1.17%** translates to **1,800+** additional license plates being misrecognized under the fair split (vs. the standard one);
- The highest gap of **5.20%** represents a staggering number of **8,000+** more license plates being incorrectly recognized under the fair split.

[4] CCPD's standard protocol: 19.1% of the test images have duplicates in the training set.

Results – Overview

The high fraction of near-duplicates in the splits traditionally adopted in the literature ***may have hindered the development and acceptance of more efficient LPR models*** *that have strong generalization abilities but do not memorize duplicates as well as other models.*

- The list of near-duplicates we have found and proposals for fair splits are publicly available for further research at https://raysonlaroca.github.io/supp/lpr-train-on-test/

A First Look at Dataset Bias in License Plate Recognition

Name that Dataset!

Can you name the dataset to which each of these images belongs?

RodoSol-ALPR (ES): ___ , ___ , ___ , ___ SSIG-SegPlate (MG): ___ , ___ , ___ , ___

UFOP (MG): ___ , ___ , ___ , ___ UFPR-ALPR (PR): ___ , ___ , ___

Name that Dataset!

RodoSol-ALPR (ES): **(a)**, **(d)**, **(h)**, **(l)**

SSIG-SegPlate (MG): **(e)**, **(i)**, **(j)**, **(o)**

UFOP (MG): **(b)**, **(f)**, **(m)**, **(n)**

UFPR-ALPR (PR): **(c)**, **(g)**, **(k)**

Name that Dataset!

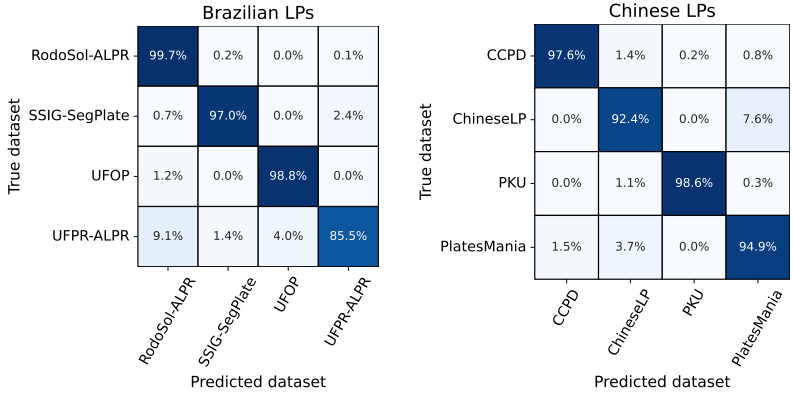
RodoSol-ALPR (ES): **(a)**, **(d)**, **(h)**, **(l)** SSIG-SegPlate (MG): **(e)**, **(i)**, **(j)**, **(o)**

UFOP (MG): **(b)**, **(f)**, **(m)**, **(n)** UFPR-ALPR (PR): **(c)**, **(g)**, **(k)**

- A shallow CNN (3 conv. layers) predicts the correct dataset in more than **95%** of cases[5].

[5] (chance is $1/4 = 25\%$)

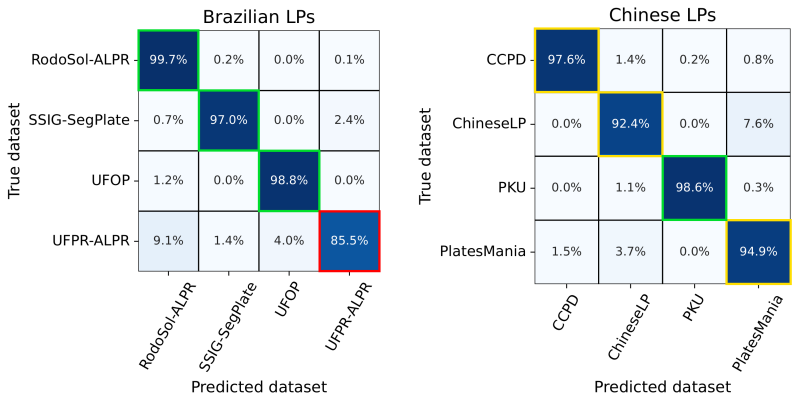
Results



There is a **clearly pronounced diagonal** in both matrices, indicating that **each dataset does have a unique, identifiable "signature."**

The overall accuracy was **95.2%** for Brazilian LPs and **95.9%** for Chinese LPs.

Results



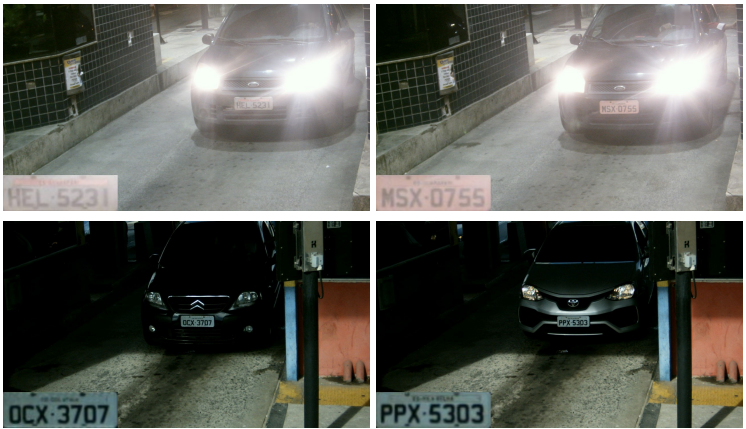
The model is more successful in classifying LP images from the datasets acquired with **static** cameras than images from the datasets captured by **handheld** or **moving** cameras.

Results

- Images from **static cameras** have many characteristics in common, not just the background.
  - These similarities are probably present to some extent in the LP regions.

Results

- Images from **static cameras** have many characteristics in common, not just the background.
  - These similarities are probably present to some extent in the LP regions.



Discussion

Most LPR models are probably learning and exploiting such signatures to improve the results achieved in seen datasets **at the cost of losing generalization capability**.

Discussion

Most LPR models are probably learning and exploiting such signatures to improve the results achieved in seen datasets **at the cost of losing generalization capability**.

SSIG-SegPlate:

- It has **563** LP images with the letter **'O'** in the first position;

Discussion

Most LPR models are probably learning and exploiting such signatures to improve the results achieved in seen datasets **at the cost of losing generalization capability**.

SSIG-SegPlate:

- It has **563** LP images with the letter **'O'** in the first position;

- It has **no** LP images with the letter **'Q'** in the first position.

Discussion

Most LPR models are probably learning and exploiting such signatures to improve the results achieved in seen datasets **at the cost of losing generalization capability**.

SSIG-SegPlate:

- It has **563** LP images with the letter **'O'** in the first position;

- It has **no** LP images with the letter **'Q'** in the first position.

Taking this into account:

- An LPR model capable of identifying that a given LP image belongs to the SSIG-SegPlate dataset **may predict the letter 'O' as the first character** even if the character looks more like 'Q' than 'O' due to noise, shadows, or other factors.

Discussion

Most LPR models are probably learning and exploiting such signatures to improve the results achieved in seen datasets **at the cost of losing generalization capability**.

SSIG-SegPlate:

- It has **563** LP images with the letter **'O'** in the first position;

- It has **no** LP images with the letter **'Q'** in the first position.

Taking this into account:

- An LPR model capable of identifying that a given LP image belongs to the SSIG-SegPlate dataset **may predict the letter 'O' as the first character** even if the character looks more like 'Q' than 'O' due to noise, shadows, or other factors.
  - However, the potentially high recognition rates achieved in the *SSIG-SegPlate* dataset would likely not be reached in unseen datasets.

Discussion

Probable causes of dataset bias in the LPR context:

- The **cameras** used to collect the images in each dataset;
- How the images were **stored** in different datasets;
  - e.g., CCPD contains highly compressed images, while most other datasets do not.

Discussion

Probable causes of dataset bias in the LPR context:

- The **cameras** used to collect the images in each dataset;
- How the images were **stored** in different datasets;
  - e.g., CCPD contains highly compressed images, while most other datasets do not.

Two initial ways to mitigate the dataset bias problem in LPR:

- Leveraging deep learning-based methods' high capability to visualize and understand how bias has crept into the datasets;
  - One technique that immediately comes to mind is Grad-CAM.
- To embrace the "wildness" of the **internet** to collect a large-scale dataset for LPR.
  - Multiple sources (e.g., multiple search engines and websites from various countries).

Conclusions

Conclusions

Automatic License Plate Recognition (ALPR): Toward Improving the State of the Art and Bridging the Gap Between Academia and Industry

RodoSol-ALPR
- Mercosur LPs
- Motorcycles & Two-row LPs

Model Fusion
- Majority Vote > Highest Conf.
- Combine 4-6 fast models for optimal speed/accuracy trade-off

Synthetic Data
- Templates, Permutation & GAN
- Synergistic effect
- Effective with limited real data

Cross-Dataset
- Major drops in performance
- Importance of RodoSol-ALPR

Near-Duplicates

Dataset Bias

Thank you!

https://raysonlaroca.github.io/