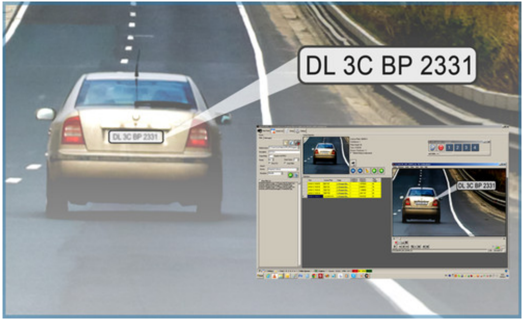
On the Cross-Dataset Generalization in License Plate Recognition

Rayson Laroca[1], Everton V. Cardoso[1], Diego R. Lucio[1],
Valter Estevam[2], David Menotti[1]

[1]Federal University of Paraná, Curitiba, Brazil
[2]Federal Institute of Paraná, Irati, Brazil

February 6, 2022

Automatic License Plate Recognition (ALPR)

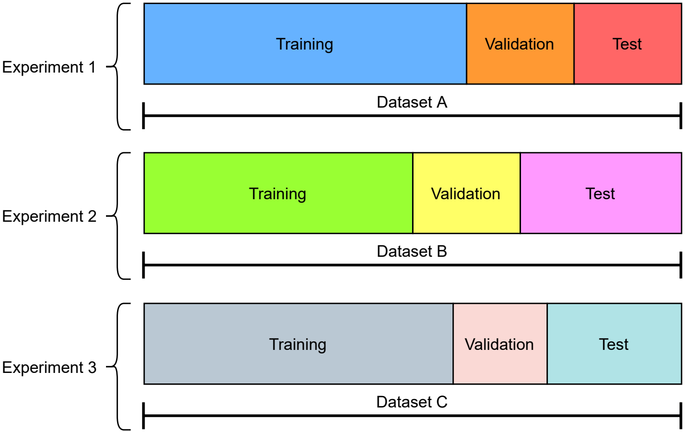


Source: Google Images

- Automatic License Plate Recognition (ALPR) aims to detect and recognize the characters on license plates (LPs) from images or videos;
- Many **practical applications** such as road traffic monitoring, toll collection, and vehicle access control in restricted areas.

Introduction - Traditional Split [1/2]

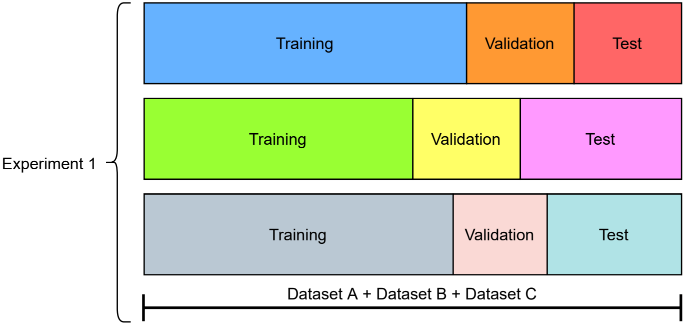
In the past, the evaluation of ALPR systems used to be done within each dataset.



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

Introduction - Traditional Split [2/2]

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



- Deep learning-based ALPR systems have often achieved **recognition rates above 99%** in several public datasets under this protocol (traditional-split).

Generalization ability?

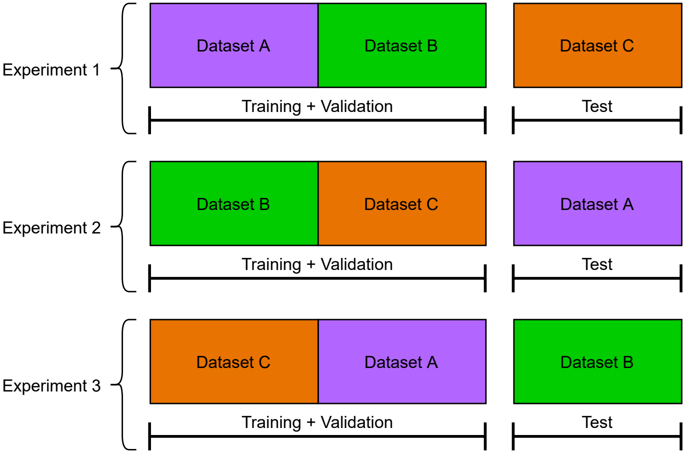
Introduction - Generalizability

Generalization ability?

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

Introduction - Leave-one-dataset-out

A **leave-one-dataset-out protocol** enables simulating this specific scenario and providing an adequate evaluation of the generalizability of the models.



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

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

Introduction - Mercosur LPs

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

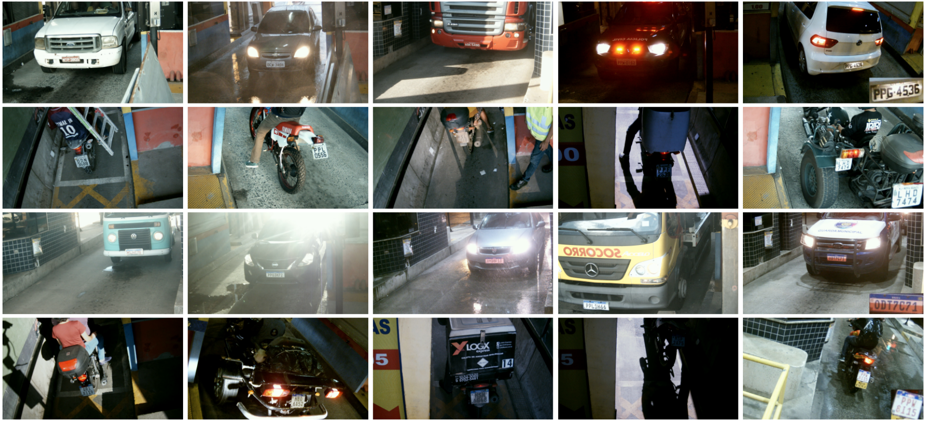
There is still no public dataset for ALPR with images of Mercosur LPs.

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

Contributions

Hence, in this paper we propose:

- A **traditional-split *versus* leave-one-dataset-out** setup that can be considered a valid testbed for cross-dataset generalization methods proposed in future works.
  - We focus on the LP recognition stage since it is the current bottleneck of ALPR systems.
- **A public dataset for ALPR**, called RodoSol-ALPR, with 20,000 images acquired in real-world scenarios, being half of them of vehicles with Mercosur LPs.



Sample images of the RodoSol-ALPR dataset, which is publicly available to the research community at https://github.com/raysonlaroca/rodosol-alpr-dataset/.

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

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).

Experimental Setup

Traditional-split *versus* leave-one-dataset-out experimental setup:

- **12** recognition models;
- RodoSol-ALPR **+ eight** well-known public datasets;
- Performance Evaluation.

Experimental Setup - Recognition Models

Table 1: OCR models explored in our experiments.

Model	Original Application
Framework: PyTorch	
R2AM (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 (Gonçalves et al., 2018)	License Plate Recognition
Framework: Darknet	
CR-NET (Silva and Jung, 2020)	License Plate Recognition
Fast-OCR (Laroca et al., 2021a)	Image-based Meter Reading

Experimental Setup - Datasets [1/2][2]

Table 2: The datasets used in our experiments.

Dataset	Year	Images	Resolution	LP Layout
Caltech Cars	1999	126	896×592	American
EnglishLP	2003	509	640×480	European
UCSD-Stills	2005	291	640×480	American
ChineseLP	2012	411	Various	Chinese
AOLP	2013	2049	Various	Taiwanese
OpenALPR-EU	2016	108	Various	European
SSIG-SegPlate	2016	2000	1920×1080	Brazilian
UFPR-ALPR	2018	4500	1920×1080	Brazilian
RodoSol-ALPR	2022	20000	1280×720	Brazilian/Mercosur

[2] In this work, the "Chinese" layout refers to LPs of vehicles registered in mainland China, while the "Taiwanese" layout refers to LPs of vehicles registered in the Taiwan region.

Experimental Setup - Datasets [2/2]

(a) Caltech Cars

(b) EnglishLP

(c) UCSD-Stills

(d) ChineseLP

(e) AOLP

(f) OpenALPR-EU

(g) SSIG-SegPlate

(h) UFPR-ALPR

Experimental Setup - Data Augmentation

To prevent overfitting and eliminate biases, we balanced the number of training images from different datasets through a series of data augmentation techniques.

Experimental Setup - Performance Evaluation

For each experiment, we report the number of correctly recognized LPs divided by the number of LPs in the test set.

- A correctly recognized LP means that **all characters** on the LP were correctly recognized.

Results - Traditional Split

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

Approach \ Test set	Caltech Cars	EnglishLP	UCSD-Stills	ChineseLP	AOLP	OpenALPR-EU	SSIG-SegPlate	UFPR-ALPR	RodoSol-ALPR	Average
CR-NET	**95.7%**	92.2%	**100.0%**	96.9%	97.7%	**97.2%**	97.1%	**78.3%**	55.8%	**90.1%**
CRNN	87.0%	81.4%	88.3%	88.2%	87.6%	89.8%	93.4%	64.9%	48.2%	81.0%
Fast-OCR	93.5%	81.4%	95.0%	85.1%	95.8%	91.7%	87.1%	65.9%	49.7%	82.8%
GRCNN	93.5%	87.3%	91.7%	84.5%	85.9%	87.0%	94.3%	63.3%	48.4%	81.7%
Holistic-CNN	89.1%	68.6%	88.3%	90.7%	86.3%	78.7%	94.8%	70.3%	49.0%	79.5%
Multi-task	87.0%	62.7%	85.0%	86.3%	84.7%	66.7%	93.0%	65.3%	49.1%	75.5%
R2AM	84.8%	70.6%	81.7%	87.0%	83.1%	63.9%	92.0%	66.9%	48.6%	75.4%
RARE	91.3%	84.3%	90.0%	95.7%	93.4%	91.7%	93.7%	69.0%	51.6%	84.5%
Rosetta	87.0%	75.5%	81.7%	90.1%	83.7%	81.5%	94.3%	63.9%	48.7%	78.5%
STAR-Net	**95.7%**	**93.1%**	96.7%	96.9%	96.8%	95.4%	96.1%	70.9%	51.8%	88.2%
TRBA	91.3%	87.3%	96.7%	96.9%	**99.0%**	93.5%	**97.3%**	72.9%	**59.6%**	88.3%
ViTSTR-Base	84.8%	80.4%	90.0%	**99.4%**	95.6%	84.3%	96.1%	73.3%	49.3%	83.7%
Average	90.0%	80.4%	90.4%	91.5%	90.8%	85.1%	94.1%	68.7%	50.8%	82.4%

Results - Traditional Split

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

Approach \ Test set	Caltech Cars	EnglishLP	UCSD-Stills	ChineseLP	AOLP	OpenALPR-EU	SSIG-SegPlate	UFPR-ALPR	RodoSol-ALPR	Average
CR-NET	**95.7%**	92.2%	**100.0%**	96.9%	97.7%	**97.2%**	97.1%	**78.3%**	55.8%	**90.1%**
CRNN	87.0%	81.4%	88.3%	88.2%	87.6%	89.8%	93.4%	64.9%	48.2%	81.0%
Fast-OCR	93.5%	81.4%	95.0%	85.1%	95.8%	91.7%	87.1%	65.9%	49.7%	82.8%
GRCNN	93.5%	87.3%	91.7%	84.5%	85.9%	87.0%	94.3%	63.3%	48.4%	81.7%
Holistic-CNN	89.1%	68.6%	88.3%	90.7%	86.3%	78.7%	94.8%	70.3%	49.0%	79.5%
Multi-task	87.0%	62.7%	85.0%	86.3%	84.7%	66.7%	93.0%	65.3%	49.1%	75.5%
R2AM	84.8%	70.6%	81.7%	87.0%	83.1%	63.9%	92.0%	66.9%	48.6%	75.4%
RARE	91.3%	84.3%	90.0%	95.7%	93.4%	91.7%	93.7%	69.0%	51.6%	84.5%
Rosetta	87.0%	75.5%	81.7%	90.1%	83.7%	81.5%	94.3%	63.9%	48.7%	78.5%
STAR-Net	**95.7%**	**93.1%**	96.7%	96.9%	96.8%	95.4%	96.1%	70.9%	51.8%	88.2%
TRBA	91.3%	87.3%	96.7%	96.9%	**99.0%**	93.5%	**97.3%**	72.9%	**59.6%**	88.3%
ViTSTR-Base	84.8%	80.4%	90.0%	**99.4%**	95.6%	84.3%	96.1%	73.3%	49.3%	83.7%
Average	90.0%	80.4%	90.4%	91.5%	90.8%	85.1%	94.1%	68.7%	50.8%	82.4%

Some datasets are not as challenging as they were when they were first proposed.

Results - Traditional Split

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

Approach \ Test set	Caltech Cars	EnglishLP	UCSD-Stills	ChineseLP	AOLP	OpenALPR-EU	SSIG-SegPlate	UFPR-ALPR	RodoSol-ALPR	Average
CR-NET	**95.7%**	92.2%	**100.0%**	96.9%	97.7%	**97.2%**	97.1%	**78.3%**	55.8%	**90.1%**
CRNN	87.0%	81.4%	88.3%	88.2%	87.6%	89.8%	93.4%	64.9%	48.2%	81.0%
Fast-OCR	93.5%	81.4%	95.0%	85.1%	95.8%	91.7%	87.1%	65.9%	49.7%	82.8%
GRCNN	93.5%	87.3%	91.7%	84.5%	85.9%	87.0%	94.3%	63.3%	48.4%	81.7%
Holistic-CNN	89.1%	68.6%	88.3%	90.7%	86.3%	78.7%	94.8%	70.3%	49.0%	79.5%
Multi-task	87.0%	62.7%	85.0%	86.3%	84.7%	66.7%	93.0%	65.3%	49.1%	75.5%
R2AM	84.8%	70.6%	81.7%	87.0%	83.1%	63.9%	92.0%	66.9%	48.6%	75.4%
RARE	91.3%	84.3%	90.0%	95.7%	93.4%	91.7%	93.7%	69.0%	51.6%	84.5%
Rosetta	87.0%	75.5%	81.7%	90.1%	83.7%	81.5%	94.3%	63.9%	48.7%	78.5%
STAR-Net	**95.7%**	**93.1%**	96.7%	96.9%	96.8%	95.4%	96.1%	70.9%	51.8%	88.2%
TRBA	91.3%	87.3%	96.7%	96.9%	**99.0%**	93.5%	**97.3%**	72.9%	**59.6%**	88.3%
ViTSTR-Base	84.8%	80.4%	90.0%	**99.4%**	95.6%	84.3%	96.1%	73.3%	49.3%	83.7%
Average	90.0%	**80.4%**	90.4%	91.5%	90.8%	85.1%	94.1%	**68.7%**	**50.8%**	82.4%

What do these datasets have in common?

Results - 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 27% of the total vehicle fleet.[3]
  - All motorcycles in Brazil have two-row LPs.

[3] www.gov.br/infraestrutura/pt-br/assuntos/transito/conteudo-denatran/frota-de-veiculos-2021

Results - Leave-one-dataset-out protocol

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

Approach \ Test set	Caltech Cars	EnglishLP	UCSD-Stills	ChineseLP	AOLP	OpenALPR-EU	SSIG-SegPlate	UFPR-ALPR	RodoSol-ALPR	Average
CR-NET	93.5%	**96.1%**	**96.7%**	88.2%	76.9%	**96.3%**	94.7%	61.8%	45.4%	**83.3%**
CRNN	91.3%	62.7%	75.0%	76.4%	59.4%	88.0%	91.3%	61.7%	38.8%	71.6%
Fast-OCR	93.5%	91.2%	95.0%	90.1%	77.0%	94.4%	91.2%	53.2%	**47.8%**	81.5%
GRCNN	**95.7%**	65.7%	90.0%	80.7%	53.9%	88.9%	90.3%	60.8%	39.8%	74.0%
Holistic-CNN	80.4%	40.2%	73.3%	81.4%	59.7%	83.3%	93.4%	61.8%	33.4%	67.4%
Multi-task	82.6%	34.3%	66.7%	77.6%	50.8%	79.6%	89.9%	57.9%	44.8%	64.9%
R2AM	89.1%	52.9%	66.7%	74.5%	52.5%	80.6%	93.5%	57.9%	40.7%	67.6%
RARE	84.8%	50.0%	85.0%	88.8%	62.9%	91.7%	93.5%	71.3%	40.1%	74.2%
Rosetta	89.1%	63.7%	68.3%	83.2%	51.1%	81.5%	94.4%	61.8%	42.5%	70.6%
STAR-Net	89.1%	80.4%	91.7%	**95.0%**	**79.3%**	93.5%	94.0%	69.1%	43.6%	81.8%
TRBA	**95.7%**	66.7%	93.3%	**95.0%**	70.0%	92.6%	96.9%	**73.2%**	42.6%	80.7%
ViTSTR-Base	89.1%	58.8%	90.0%	**95.0%**	59.2%	89.8%	**97.9%**	69.6%	41.7%	76.8%
Average	89.5%	63.6%	82.6%	85.5%	62.7%	88.3%	93.4%	63.3%	41.8%	74.5%
Avg. (traditional-split)	90.0%	80.4%	90.4%	91.5%	90.8%	85.1%	94.1%	68.7%	50.8%	82.4%
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%

Results - Leave-one-dataset-out protocol

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

Approach \ Test set	Caltech Cars	EnglishLP	UCSD-Stills	ChineseLP	AOLP	OpenALPR-EU	SSIG-SegPlate	UFPR-ALPR	RodoSol-ALPR	Average
CR-NET	93.5%	**96.1%**	**96.7%**	88.2%	76.9%	**96.3%**	94.7%	61.8%	45.4%	**83.3%**
CRNN	91.3%	62.7%	75.0%	76.4%	59.4%	88.0%	91.3%	61.7%	38.8%	71.6%
Fast-OCR	93.5%	91.2%	95.0%	90.1%	77.0%	94.4%	91.2%	53.2%	**47.8%**	81.5%
GRCNN	**95.7%**	65.7%	90.0%	80.7%	53.9%	88.9%	90.3%	60.8%	39.8%	74.0%
Holistic-CNN	80.4%	40.2%	73.3%	81.4%	59.7%	83.3%	93.4%	61.8%	33.4%	67.4%
Multi-task	82.6%	34.3%	66.7%	77.6%	50.8%	79.6%	89.9%	57.9%	44.8%	64.9%
R2AM	89.1%	52.9%	66.7%	74.5%	52.5%	80.6%	93.5%	57.9%	40.7%	67.6%
RARE	84.8%	50.0%	85.0%	88.8%	62.9%	91.7%	93.5%	71.3%	40.1%	74.2%
Rosetta	89.1%	63.7%	68.3%	83.2%	51.1%	81.5%	94.4%	61.8%	42.5%	70.6%
STAR-Net	89.1%	80.4%	91.7%	**95.0%**	**79.3%**	93.5%	94.0%	69.1%	43.6%	81.8%
TRBA	**95.7%**	66.7%	93.3%	**95.0%**	70.0%	92.6%	96.9%	**73.2%**	42.6%	80.7%
ViTSTR-Base	89.1%	58.8%	90.0%	**95.0%**	59.2%	89.8%	**97.9%**	69.6%	41.7%	76.8%
Average	89.5%	**63.6%**	82.6%	85.5%	**62.7%**	88.3%	93.4%	63.3%	41.8%	**74.5%**
Avg. (traditional-split)	90.0%	80.4%	90.4%	91.5%	90.8%	85.1%	94.1%	68.7%	50.8%	82.4%
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%

Results - Leave-one-dataset-out protocol

LODO: `7615BG`
Trad.: `7615RG`

LODO: `PG379I`
Trad.: `P63791`

LODO: `0X7655`
Trad.: `DX7655`

The predictions obtained by TRBA (Baek et al., 2019) on three images of the AOLP dataset.

LODO: `NB4071P`
Trad.: `MB4071P`

LODO: `ZG56-2AB`
Trad.: `ZG5642AB`

LODO: `BJ28GBM`
Trad.: `BJ286BM`

The predictions obtained by STAR-Net (Liu et al., 2016) on three images of the EnglishLP dataset.

In general, the errors (outlined in red) 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 - Leave-one-dataset-out protocol

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

Approach \ Test set	Caltech Cars	EnglishLP	UCSD-Stills	ChineseLP	AOLP	OpenALPR-EU	SSIG-SegPlate	UFPR-ALPR	RodoSol-ALPR	Average
CR-NET	93.5%	**96.1%**	**96.7%**	88.2%	76.9%	**96.3%**	94.7%	61.8%	45.4%	**83.3%**
CRNN	91.3%	62.7%	75.0%	76.4%	59.4%	88.0%	91.3%	61.7%	38.8%	71.6%
Fast-OCR	93.5%	91.2%	95.0%	90.1%	77.0%	94.4%	91.2%	53.2%	**47.8%**	81.5%
GRCNN	**95.7%**	65.7%	90.0%	80.7%	53.9%	88.9%	90.3%	60.8%	39.8%	74.0%
Holistic-CNN	80.4%	40.2%	73.3%	81.4%	59.7%	83.3%	93.4%	61.8%	33.4%	67.4%
Multi-task	82.6%	34.3%	66.7%	77.6%	50.8%	79.6%	89.9%	57.9%	44.8%	64.9%
R2AM	89.1%	52.9%	66.7%	74.5%	52.5%	80.6%	93.5%	57.9%	40.7%	67.6%
RARE	84.8%	50.0%	85.0%	88.8%	62.9%	91.7%	93.5%	71.3%	40.1%	74.2%
Rosetta	89.1%	63.7%	68.3%	83.2%	51.1%	81.5%	94.4%	61.8%	42.5%	70.6%
STAR-Net	89.1%	80.4%	91.7%	**95.0%**	**79.3%**	93.5%	94.0%	69.1%	43.6%	81.8%
TRBA	**95.7%**	66.7%	93.3%	**95.0%**	70.0%	92.6%	96.9%	**73.2%**	42.6%	80.7%
ViTSTR-Base	89.1%	58.8%	90.0%	**95.0%**	59.2%	89.8%	**97.9%**	69.6%	41.7%	76.8%
Average	89.5%	63.6%	82.6%	85.5%	62.7%	88.3%	93.4%	63.3%	41.8%	74.5%
Avg. (traditional-split)	90.0%	80.4%	90.4%	91.5%	90.8%	85.1%	94.1%	68.7%	50.8%	82.4%
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%

Six different models obtained the best result in at least one dataset.

Results - Leave-one-dataset-out protocol

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

Approach \ Test set	Caltech Cars	EnglishLP	UCSD-Stills	ChineseLP	AOLP	OpenALPR-EU	SSIG-SegPlate	UFPR-ALPR	RodoSol-ALPR	Average
CR-NET	93.5%	**96.1%**	**96.7%**	88.2%	76.9%	**96.3%**	94.7%	61.8%	45.4%	**83.3%**
CRNN	91.3%	62.7%	75.0%	76.4%	59.4%	88.0%	91.3%	61.7%	38.8%	71.6%
Fast-OCR	93.5%	91.2%	95.0%	90.1%	77.0%	94.4%	91.2%	53.2%	**47.8%**	81.5%
GRCNN	**95.7%**	65.7%	90.0%	80.7%	53.9%	88.9%	90.3%	60.8%	39.8%	74.0%
Holistic-CNN	80.4%	40.2%	73.3%	81.4%	59.7%	83.3%	93.4%	61.8%	33.4%	67.4%
Multi-task	82.6%	34.3%	66.7%	77.6%	50.8%	79.6%	89.9%	57.9%	44.8%	64.9%
R2AM	89.1%	52.9%	66.7%	74.5%	52.5%	80.6%	93.5%	57.9%	40.7%	67.6%
RARE	84.8%	50.0%	85.0%	88.8%	62.9%	91.7%	93.5%	71.3%	40.1%	74.2%
Rosetta	89.1%	63.7%	68.3%	83.2%	51.1%	81.5%	94.4%	61.8%	42.5%	70.6%
STAR-Net	89.1%	80.4%	91.7%	**95.0%**	**79.3%**	93.5%	94.0%	69.1%	43.6%	81.8%
TRBA	**95.7%**	66.7%	93.3%	**95.0%**	70.0%	92.6%	96.9%	**73.2%**	42.6%	80.7%
ViTSTR-Base	89.1%	58.8%	90.0%	**95.0%**	59.2%	89.8%	**97.9%**	69.6%	41.7%	76.8%
Average	89.5%	63.6%	82.6%	85.5%	62.7%	88.3%	93.4%	63.3%	**41.8%**	74.5%
Avg. (traditional-split)	90.0%	80.4%	90.4%	91.5%	90.8%	85.1%	94.1%	68.7%	**50.8%**	82.4%
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%

The RodoSol-ALPR dataset proved very challenging.

Results - Leave-one-dataset-out protocol[4]

TRBA: HLPA594
OpenALPR: HLP4594
GT: HLP4594

TRBA: PPY6O26
OpenALPR: PPY6O26
GT: PPY6C26

TRBA: QRE4E6Z
OpenALPR: QRE4E62
GT: QRE4E62

TRBA: QRG6D57
OpenALPR: -------
GT: QRG6D57

TRBA: OOM8060
OpenALPR: OOM8060
GT: ODM8060

TRBA: MRO3095
OpenALPR: MRO3095
GT: MRU3095

Some LP images from RodoSol-ALPR along with the predictions returned by TRBA and OpenALPR.

[4] For correctness, we checked if the ground truth (GT) matched the vehicle make and model on the National Traffic Department of Brazil (DENATRAN) database.

Conclusions

- Researchers should pay more attention to **cross-dataset LP recognition**;
  - There are significant drops in performance (e.g., 90.8% $\rightarrow$ 62.7%) for most datasets when training and testing the recognition models in a leave-one-dataset-out fashion;
  - It better simulates real-world ALPR applications, where new cameras are regularly being installed in new locations without existing systems being retrained as often.

Conclusions

- Researchers should pay more attention to **cross-dataset LP recognition**;
  - There are significant drops in performance (e.g., 90.8% $\rightarrow$ 62.7%) for most datasets when training and testing the recognition models in a leave-one-dataset-out fashion;
  - It better simulates real-world ALPR applications, where new cameras are regularly being installed in new locations without existing systems being retrained as often.
- It is important to perform experiments on **multiple datasets**;
  - Six different models reached the best result in at least one dataset under the leave-one-dataset-out protocol.

Conclusions

- Researchers should pay more attention to **cross-dataset LP recognition**;
  - There are significant drops in performance (e.g., 90.8% $\rightarrow$ 62.7%) for most datasets when training and testing the recognition models in a leave-one-dataset-out fashion;
  - It better simulates real-world ALPR applications, where new cameras are regularly being installed in new locations without existing systems being retrained as often.
- It is important to perform experiments on **multiple datasets**;
  - Six different models reached the best result in at least one dataset under the leave-one-dataset-out protocol.
- The **RodoSol-ALPR dataset** has proved challenging.
  - Both the models trained by us and two well-known commercial systems reached recognition rates below 70% on its test set;
  - It will assist in developing new approaches for Mercosur LPs (including two-row ones) and the fair comparison between methods proposed in different works.

Thank you!

https://raysonlaroca.github.io/