LPLC: A Dataset for License Plate Legibility Classification

Lucas Wojcik[*], Gabriel E. Lima[*], Valfride Nascimento[*],
Eduil Nascimento Jr.[†], Rayson Laroca[‡,*], David Menotti[*]

[*] Department of Informatics, Federal University of Paraná, Curitiba, Brazil
[†] Department of Technological Development and Quality, Paraná Military Police, Curitiba, Brazil
[‡] Graduate Program in Informatics, Pontifical Catholic University of Paraná, Curitiba, Brazil

Motivation

- ALPR remains unsolved for **challenging** scenarios
  - Adversarial environments, poor equipment, transmission compression

- **Super-resolution** (SR) presents new challenges
  - Good images get degraded: which images should undergo SR?

- Real-world systems deal with untreatable **large amounts of data**
  - How much of it can or should be discarded?

Quality vs. Legibility

Good quality images (top right) may feature illegible LPs and vice-versa (bottom right)
A single image may contain both legible and illegible LPs.

Legibility Example

CAPES
UFPR
UNIVERSIDADE FEDERAL DO PARANÁ
PUCPR
GRUPO MARISTA

The Dataset

- Real-world radar images from the Brazilian state of Paraná

- Annotated attributes:
  - 4-corner bounding box
  - Plate OCR (characters)
  - Legibility class (4 levels)
  - Vehicle and plate occlusion flags

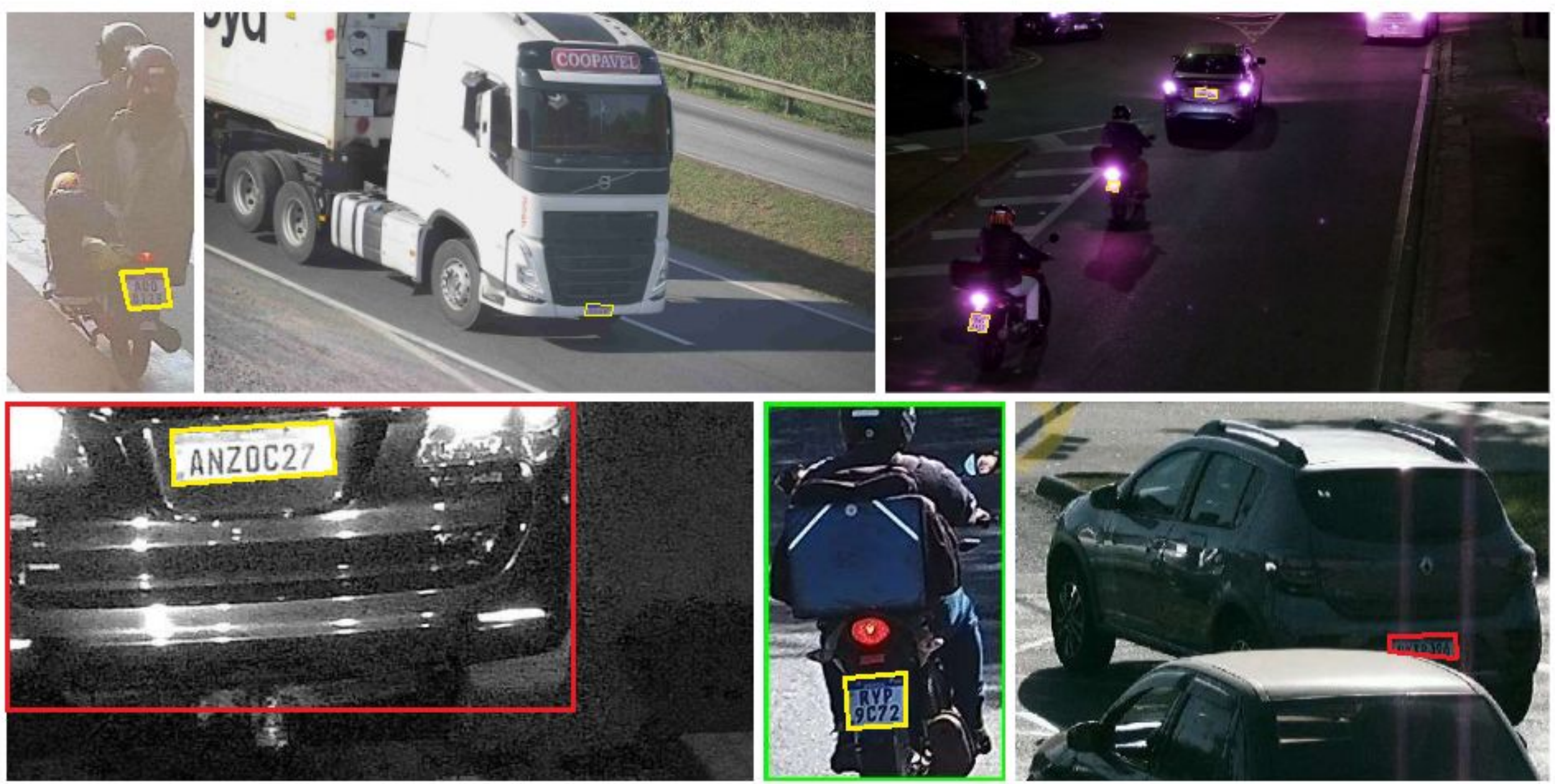
- For for various ALPR tasks outside of Legibility Classification

CAPES UFPR UNIVERSIDADE FEDERAL DO PARANÁ PARANÁ POLÍCIA MILITAR SCIENTIA, VITA ET FIDES PUCPR GRUPO MARISTA

The Dataset – Instances and Attributes

The Dataset – Instances and Attributes

The Dataset – Instances and Attributes



Legibility Levels

Perfect | Good | Poor | Illegible

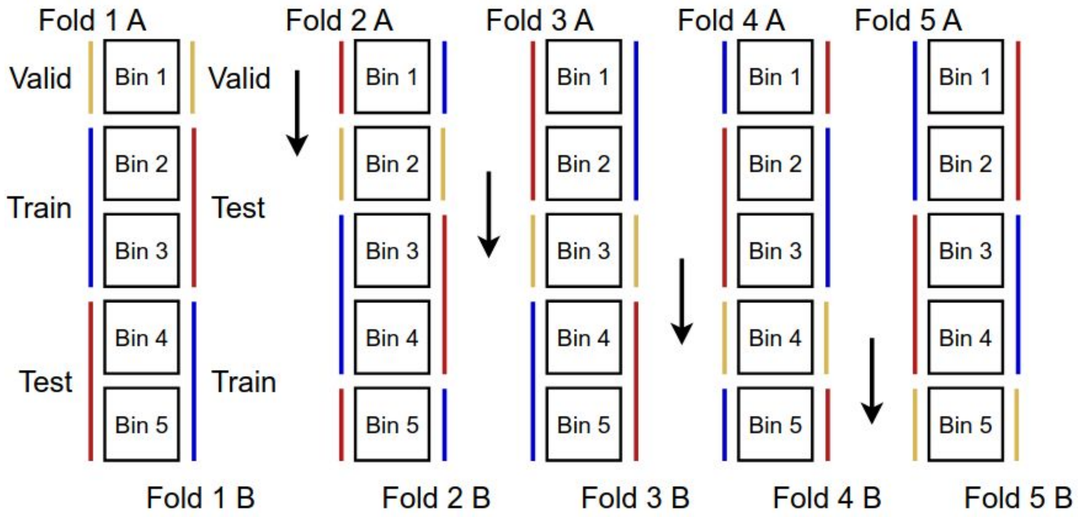
CAPES
UFPR
UNIVERSIDADE FEDERAL DO PARANÁ
PUCPR
GRUPO MARISTA

Legibility Levels

LPs by Legibility		Other Attributes		
Class	Number	Class	True	False
Perfect	5,617	Occluded	12,586	101
Good	3,641	Valid	12,359	328
Poor	1,825	OCR	11,083	1,604
Illegible	1,604	→ Total LPs	12,687	

Images by Time of Day		Images by Attributes	
Class	Number	Has at Least One	Number
Morning	3,830	Legible LPs	9,684
Afternoon	2,556	Non Occluded LPs	10,195
Evening	2,585	Valid Vehicles	10,030
Night	1,239	→ Total Images	10,200

Legibility Classification



Our 10-fold split protocol

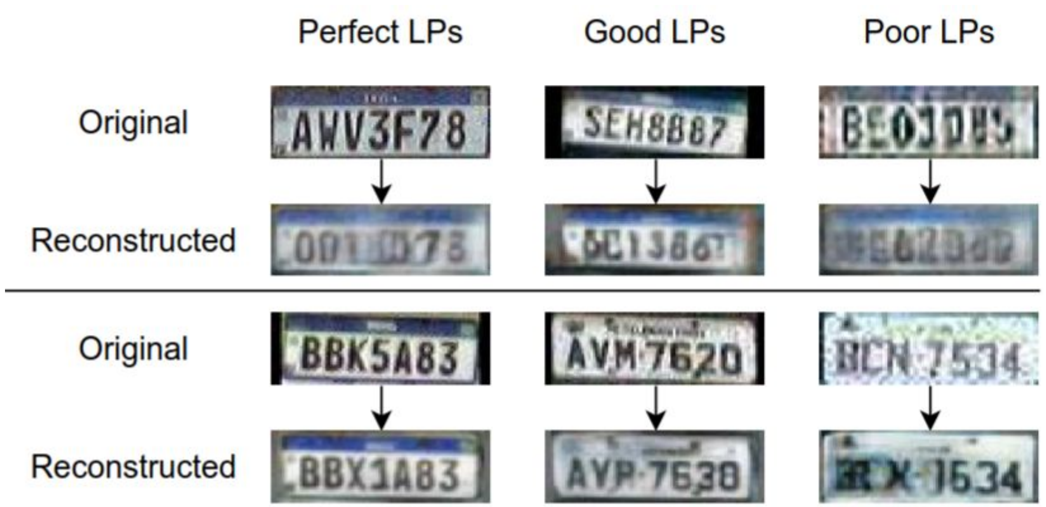
Model	Class				Overall
	Perfect	Good	Poor	Illegible	
ResNet-50	84.54%	67.98%	56.70%	72.97%	74.51%
ViT b-16	85.74%	68.00%	58.80%	73.67%	75.48%
YOLO11m-cls	88.37%	65.83%	59.42%	74.47%	76.79%

Model	Legibility Recognition (Legible vs. Poor)	Full Recognition (Legible, Poor, Illegible)
ResNet-50	92.56%	87.23%
ViT b-16	93.16%	87.78%
YOLO11m-cls	92.71%	86.25%

Average test Micro-F1 score for the three legibility protocols

Super-resolution

- Out-of-the-box SR showcases poor cross-dataset performance



GAN Model	OCR Results With SR	Class			Total
		Perfect	Good	Poor	
LCOFL-GAN [9]	Better	98	108	108	314
	Equal	494	308	170	972
	Worse	5,012	3,206	1,523	9,741
Real-ESRGAN [15]	Better	211	276	160	647
	Equal	5,180	2,683	572	8,435
	Worse	213	663	1,069	1,945
LPSRGAN [10]	Better	108	98	45	251
	Equal	264	209	115	588
	Worse	5,232	3,315	1,641	10,188

Future Work

- A proper ALPR SOTA comparison
  - In traditional tasks
- Better SR methods for a cross-dataset scenario

Thank you for your attention!

Please visit us at: github.com/lmlwojcik/lplc-dataset

Our dataset is public for **research** purposes

GitHub