Leveraging Model Fusion for Improved License Plate Recognition

Rayson Laroca[1], Luiz A. Zanlorensi[1], Valter Estevam[1,2],
Rodrigo Minetto[3], and David Menotti[1]

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

Automatic License Plate Recognition (ALPR)

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

Automatic License Plate Recognition (ALPR)

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

ALPR has many practical applications:

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

A usual *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

*Recent studies have shown that recognition models demonstrate varying levels of robustness across different datasets. As each dataset poses distinct challenges, such as diverse license plate (LP) layouts and resolution, a model that performs **optimally on one dataset** often yields **poor results on another**.*

Research Questions

Can we substantially enhance LPR results across various datasets ***by fusing the outputs of multiple models?***

Research Questions

Can we substantially enhance LPR results across various datasets ***by fusing the outputs of multiple models?***

Additional Questions:

- To what extent can this enhancement be attained?
- How many and which recognition models should be employed?

Experimental Setup

- 3 primary fusion approaches.
- 12 recognition models;
- 12 public datasets;

Experimental Setup – Recognition Models

Recognition models explored in our experiments.

Model	Original Application
Framework: PyTorch[1]	
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[2]	
Holistic-CNN (Špaňhel et al., 2017)	License Plate Recognition
Multi-Task-LR (Gonçalves et al., 2019)	License Plate Recognition
Framework: Darknet[3]	
CR-NET (Silva and Jung, 2020)	License Plate Recognition
Fast-OCR (Laroca et al., 2021a)	Image-based Meter Reading

[1] https://github.com/roatienza/deep-text-recognition-benchmark/
[2] https://keras.io/
[3] https://github.com/AlexeyAB/darknet

Experimental Setup – Datasets [1/2]

There is no publicly available dataset comprising images captured from multiple regions.

- Researchers aiming to demonstrate the effectiveness of their systems for LPs from various regions must conduct experiments on multiple datasets.

Experimental Setup – Datasets [1/2]

There is no publicly available dataset comprising images captured from multiple regions.

- Researchers aiming to demonstrate the effectiveness of their systems for LPs from various regions must conduct experiments on multiple datasets.

The 12 public datasets used in our experiments.

Dataset	**Year**	**Images**	**LP Layout**
Caltech Cars	1999	126	American
EnglishLP	2003	509	European
UCSD-Stills	2005	291	American
ChineseLP	2012	411	Chinese
AOLP	2013	2,049	Taiwanese
OpenALPR-EU*	2016	108	European
SSIG-SegPlate	2016	2,000	Brazilian
PKU*	2017	2,253	Chinese
UFPR-ALPR	2018	4,500	Brazilian
CD-HARD*	2018	104	Various
CLPD*	2021	1,200	Chinese
RodoSol-ALPR	2022	20,000	Brazilian & Mercosur

*Datasets used only for testing the models (cross-dataset)

Experimental Setup – Datasets [2/2]

(a) Caltech Cars

(b) EnglishLP

(c) UCSD-Stills

(d) ChineseLP

(e) AOLP

(f) OpenALPR-EU

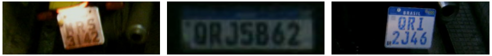
(g) SSIG-SegPlate

(h) PKU

(i) UFPR-ALPR

(j) CD-HARD

(k) CLPD



(l) RodoSol-ALPR

Some LP images from the public datasets used in our experimental evaluation.

Experimental Setup – Fusion Approaches [1/2]

Three primary approaches:

1. Highest Confidence (HC);
   - The final prediction is the sequence predicted with the highest confidence value, *even if only one model predicts it*.
2. Majority Vote (MV);
   - The final prediction is the sequence predicted by the largest number of models, *disregarding the confidence values associated with each prediction*.
3. Majority Vote by Character Position (MVCP).
   - Follows a similar MV rule but performs individual aggregation for each character position;
   - The characters predicted by all models for every position on the LP are analyzed separately. For each position, the character predicted by the largest number of models is selected. Then, the selected characters are concatenated to form the final string.

Experimental Setup – Fusion Approaches [2/2]

One concern that arises when employing majority vote-based approaches is the potential occurrence of a tie.

Two **tie-breaking approaches** for each majority vote strategy:

- Selecting the prediction made with the highest confidence among the tied predictions;
- Selecting the prediction made by the "best model".
  - For simplicity, we consider the best model the one that performs best across all datasets;
  - In a more practical scenario, the chosen model could be the one known to perform best in the specific implementation scenario (e.g., tilted LPs vs low-resolution LPs).

Results

Comparison of the recognition rates achieved across eight popular datasets by the 12 models individually and through five different fusion strategies. Each model (rows) was trained once on the combined set of training images from all datasets and evaluated on the respective test sets (columns).

Approach \ 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 the 12 models individually and through five different fusion strategies. Each model (rows) was trained once on the combined set of training images from all datasets and evaluated on the respective test sets (columns).

Approach \ 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 the 12 models individually and through five different fusion strategies. Each model (rows) was trained once on the combined set of training images from all datasets and evaluated on the respective test sets (columns).

Approach \ 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.

Approach \ 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

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

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 models were ranked based on their mean performance across the datasets.
  - The rankings on the validation and test sets were essentially the same.

Results

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

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 recognition models.

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) consistently led to worse results.
  - All models tend to make incorrect predictions also with high confidence.

Results (Qualitative)

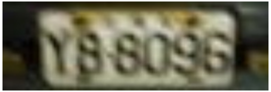
ViTSTR-Base: AIQ1Q56 (0.93)
STAR-Net: ATQ1056 (0.59)
TRBA: AIQ1056 (0.98)
CR-NET: AIQ1056 (0.82)
RARE: AIQ1Q56 (0.92)
Fusion MV-HC: AIQ1056

ViTSTR-Base: AS5I8D (0.53)
STAR-Net: AS5180 (0.82)
TRBA: AS5180 (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: MRO3095 (0.98)
TRBA: MRD3095 (0.72)
CR-NET: MRD3095 (0.94)
RARE: MRD3095 (0.87)
Fusion MV-HC: MRD3095

Predictions obtained in eight LP images using multiple models individually and the best fusion approach. The confidence for each prediction is indicated in parentheses, and any errors are highlighted in red.

Results (Qualitative)

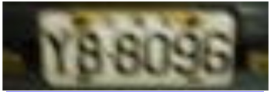
ViTSTR-Base: AIQ1Q56 (0.93)
STAR-Net: ATQ1056 (0.59)
TRBA: AIQ1056 (0.98)
CR-NET: AIQ1056 (0.82)
RARE: AIQ1Q56 (0.92)
Fusion MV-HC: AIQ1056

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: MRO3095 (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 (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

- If attaining the utmost recognition rate across various scenarios is not imperative, it becomes more advantageous to combine fewer but faster models.
  - Combining **4–6 fast models** appears to be the optimal choice for striking a better balance between speed and accuracy.

Conclusions

- First study thoroughly examining the potential improvements in LPR results across diverse datasets by fusing the outputs from multiple recognition models;
  - 12 recognition models;
  - 12 public datasets;

Conclusions

- First study thoroughly examining the potential improvements in LPR results across diverse datasets by fusing the outputs from multiple recognition models;
  - 12 recognition models;
  - 12 public datasets;

- Substantial benefits of fusion approaches in both intra- and cross-dataset setups;
  - Intra-dataset: 92.4% $\rightarrow$ 97.6% || Cross-dataset: 87.6% $\rightarrow$ 90.3%;
  - The optimal fusion approach in both setups was via **a majority vote at the sequence level**;
  - Essentially, fusing multiple models considerably reduces the likelihood of obtaining subpar performance on a particular dataset/scenario.

Conclusions

- First study thoroughly examining the potential improvements in LPR results across diverse datasets by fusing the outputs from multiple recognition models;
  - 12 recognition models;
  - 12 public datasets;
- Substantial benefits of fusion approaches in both intra- and cross-dataset setups;
  - Intra-dataset: 92.4% $\rightarrow$ 97.6% || Cross-dataset: 87.6% $\rightarrow$ 90.3%;
  - The optimal fusion approach in both setups was via **a majority vote at the sequence level**;
  - Essentially, fusing multiple models considerably reduces the likelihood of obtaining subpar performance on a particular dataset/scenario.
- 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.

CIARP 2023

Thank you!

https://raysonlaroca.github.io/supp/lpr-model-fusion/

CAPES CNPq NVIDIA UFPR UNIVERSIDADE FEDERAL DO PARANÁ