Revisiting Vehicle Color Recognition in Long-Tailed Surveillance Scenarios

Vinícius Orrú[1,2], Bruno H. Foggiatto[1], Gabriel E. Lima[3], David Menotti[3], and Rayson Laroca[1,3]

[1]Pontifical Catholic University of Paraná, Curitiba, Brazil
[2]National High Court of Brazil, Brasília, Brazil
[3]Federal University of Paraná, Curitiba, Brazil

Introduction – Vehicle Color Recognition (VCR)

- Vehicle identification is often associated with ALPR, but **license plates may be illegible** due to occlusion, motion blur, poor illumination, or low resolution.


AVV3244


AEQ1B88


ANF2072

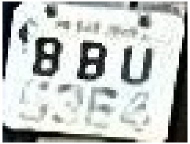

BDB8C98

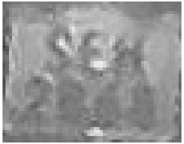

IMY2G52


SFD3H46


ASA4083


IKR5H84


BBU5364


SEH2B68

Introduction – Vehicle Color Recognition (VCR)

- **The Alternative** → **Vehicle color** is a highly practical complementary attribute:
  - Intuitive for human operators.
  - Covers a larger image region than the license plate.
  - Supports searches even when the license plate cannot be recognized.

White

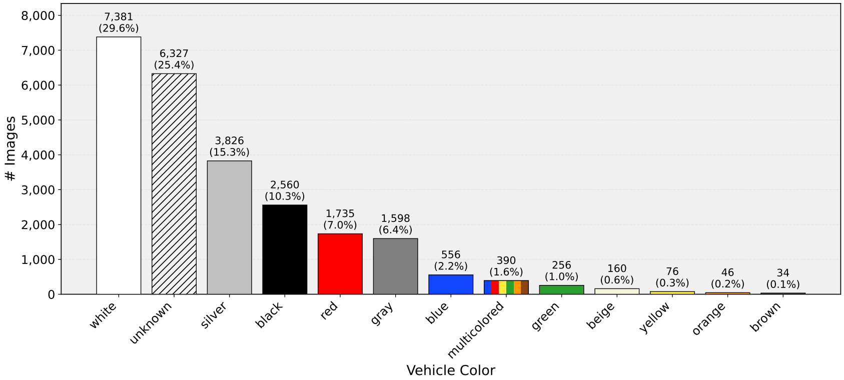
Black

Red

Blue

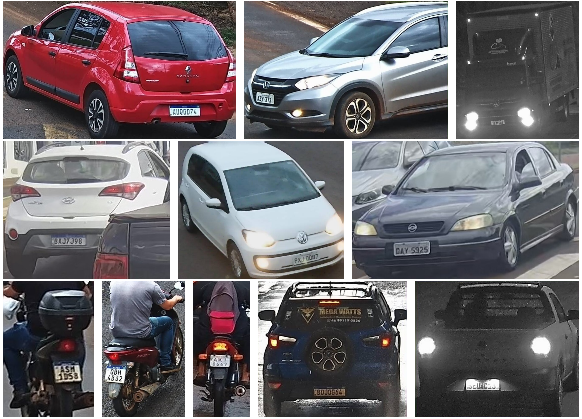
Introduction – The Long-Tailed Challenge

- **A major challenge:** Severe class imbalance naturally found in vehicle color distributions.
  - Neutral colors (white, black, silver, gray) are far more frequent than colors such as yellow, orange, green, and brown.



Dataset – UFPR-VeSV & Class Imbalance

- We use **UFPR-VeSV**[1], a highly diverse real-world surveillance dataset.
  - It contains 24,945 images of 16,297 unique vehicles across 13 color classes.



[1] G. E. Lima, V. Nascimento, E. Santos, E. Nascimento Jr., R. Laroca, D. Menotti, "Toward Unified Fine-Grained Vehicle Classification and Automatic License Plate Recognition," Journal of the Brazilian Computer Society, vol. 32, no. 1, pp. 783-799, 2026.

Proposed Solution – Synthetic Augmentation

- **Goal:** Increase the number and diversity of training samples for underrepresented colors without replacing real images.
- **Quality Control:** All generated images underwent strict manual assessment for visual realism, color correctness, and domain consistency.

Proposed Solution – Synthetic Augmentation

- **Goal:** Increase the number and diversity of training samples for underrepresented colors without replacing real images.
- **Quality Control:** All generated images underwent strict manual assessment for visual realism, color correctness, and domain consistency.
- **Strategy 1: Text-to-Image Generation**
  - **RunDiffusion/Juggernaut-XL** (12,951 accepted images).

Proposed Solution – Synthetic Augmentation

- **Goal:** Increase the number and diversity of training samples for underrepresented colors without replacing real images.
- **Quality Control:** All generated images underwent strict manual assessment for visual realism, color correctness, and domain consistency.
- **Strategy 1: Text-to-Image Generation**
  - **RunDiffusion/Juggernaut-XL** (12,951 accepted images).
- **Strategy 2: Image-Conditioned Color Editing**
  - **Gemini 2.0 Flash** (6,028 accepted images).

Synthetic Generation – Text-to-Image (RunDiffusion)

- Generates diverse vehicle models, scenes, viewpoints, and lighting conditions based on text prompts favoring minority colors.
  - **Example Prompt:** "A real **Mitsubishi Pajero**, fully in **green** color in a brazilian town, sunny day, slightly front-view".

Synthetic Generation – Text-to-Image (RunDiffusion)

- A total of 19,144 images were generated using the text-to-image model. Of these, **12,951 images** (67.7%) met the criteria for acceptance.

Synthetic Generation – Image-Conditioned Editing (Gemini)

- Change the vehicle body color while generally preserving the original camera geometry, vehicle pose, background, and image quality.
  - **Example Prompt:** "Change the color of the entire body of the vehicle in the image to **brown**, avoid vibrant tones and stick to realistic and common vehicle colors. Only change the color of the car, do not modify other elements of the image".

Original

→

Generated (Brown)

Synthetic Generation – Image-Conditioned Editing (Gemini)

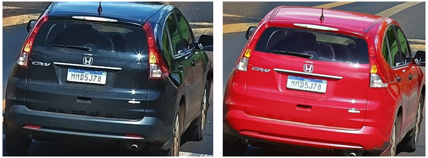
- A total of 14,040 images were generated using image-conditioned editing. Of these, **6,028 images** (42.9%) met the criteria for acceptance

Original → Red

Original → Orange

Original → Brown

Original → Yellow



Original → Red

Methodology – Representation & Optimization

- **Visual Representation:** Evaluated **frozen DINOv3 features** (Small, Base, Large variants) using a two-layer MLP head, motivated by transferable representations under limited supervision.
- **Supervised baselines:** EfficientNet-V2, ResNet-50, Swin-T, and ViT-B/16.

Methodology – Representation & Optimization

- **Visual Representation:** Evaluated **frozen DINOv3 features** (Small, Base, Large variants) using a two-layer MLP head, motivated by transferable representations under limited supervision.
- **Supervised baselines:** EfficientNet-V2, ResNet-50, Swin-T, and ViT-B/16.
- **Handling Class Imbalance**
  - Compared standard Cross-Entropy (CE) with **Weighted Cross-Entropy (WCE)**.
  - Explored **Linear Warmup with Cosine Decay (LWCD)** to stabilize early training steps.

Methodology – Augmentation & Preprocessing

- **Color-Safe Augmentation:**
  - Data augmentation must be conservative when color is the target attribute.
  - Strict limits on HSV shifts (e.g., Hue limited to ± 5 levels) to avoid corrupting the color signal.

Original

Crop

Rotate

Flip

Blur

Jitter

- **Foreground-Aware Preprocessing:**
  - Utilized **SAM 2** to preserve the vehicle region while blurring the background.
  - Assesses whether emphasizing vehicle paint reduces dependence on contextual cues.

Results – Overall Performance

- The proposed approach utilizes Gemini-augmented training data, WCE, LWCD, color-safe augmentation, SAM 2 preprocessing, and **hard-voting ensemble fusion**.

Model	Mi-Acc (%)	**Ma-Acc (%)**	F1 (%)
EfficientNet-V2 (from Lima et al.)	93.5 ± 0.6	71.5 ± 2.3	73.8 ± 2.3
DINOv3-Small	92.5 ± 0.6	69.1 ± 3.1	71.2 ± 2.2
DINOv3-Base	93.0 ± 0.4	69.6 ± 3.4	72.3 ± 3.1
DINOv3-Large	94.0 ± 0.5	72.5 ± 3.2	74.4 ± 2.5
Proposed approach	$\mathbf{94.6 \pm 0.4}$	$\mathbf{79.7 \pm 4.0}$	$\mathbf{78.8 \pm 3.1}$

- **The Impact** $\rightarrow$ Significant improvement of **8.2 percentage points** in macro-accuracy over the previous benchmark.

Results – Impact of Synthetic Data & Loss

- Isolating the effect of training-set composition:

Training configuration	Mi-Acc (%)	**Ma-Acc (%)**	F1 (%)
Cross-Entropy (CE)			
UFPR-VeSV	94.0 ± 0.5	72.5 ± 3.2	74.4 ± 2.5
UFPR-VeSV + RunDiffusion	93.5 ± 0.5	72.5 ± 4.6	74.6 ± 4.4
UFPR-VeSV + Gemini	**93.8 ± 0.3**	**74.9 ± 4.7**	**76.4 ± 3.5**
UFPR-VeSV + Gemini + RunDiffusion	93.1 ± 0.7	72.8 ± 3.8	74.1 ± 2.8
Weighted Cross-Entropy (WCE)			
UFPR-VeSV	36.4 ± 0.3	**59.3 ± 3.0**	55.8 ± 3.3
UFPR-VeSV + RunDiffusion	91.5 ± 0.4	71.5 ± 5.6	72.7 ± 5.4
UFPR-VeSV + Gemini	**92.7 ± 0.9**	**75.7 ± 3.0**	**75.0 ± 2.7**
UFPR-VeSV + Gemini + RunDiffusion	92.6 ± 0.1	71.9 ± 4.4	73.5 ± 3.5

Results – Cumulative Ablation

- The gains provided by our methodology are cumulative.

Approach	Configuration	Best model	Mi-Acc (%)	**Ma-Acc (%)**	F1 (%)	**ΔMa**
Lima et al.	Original	EfficientNet-V2	93.5 ± 0.6	71.5 ± 2.3	73.8 ± 2.3	–
Our baseline	Original + CD	DINOv3-Large	94.0 ± 0.5	72.5 ± 3.2	74.4 ± 2.5	–
+ Gemini data	Gemini + CD	DINOv3-Large	93.8 ± 0.3	74.9 ± 4.7	76.4 ± 3.5	**+2.4**
+ Loss reweighting	Gemini + WCE + CD	Swin-T	92.7 ± 0.9	75.7 ± 3.0	75.0 ± 2.7	**+0.8**
+ Scheduler	Gemini + WCE + LWCD	DINOv3-Large	93.2 ± 0.8	76.7 ± 4.3	75.3 ± 3.0	**+1.0**
+ Color-safe aug.	+ augmentation	DINOv3-Large	92.9 ± 0.5	77.3 ± 4.0	74.7 ± 2.6	**+0.6**
+ Segmented bg.	+ background blur	DINOv3-Large	93.0 ± 0.8	78.1 ± 4.6	76.0 ± 3.0	**+0.8**
+ Ensemble	hard voting	Ensemble	$\mathbf{94.6} \pm 0.4$	$\mathbf{79.7} \pm 4.0$	$\mathbf{78.8} \pm 3.1$	**+1.6**

Error Analysis – Model vs. Ambiguity

- We manually analyzed all 785 errors from three representative folds (worst, median, and best).

Error Analysis – Model vs. Ambiguity

- We manually analyzed all 785 errors from three representative folds (worst, median, and best).
- **Findings (Based on Annotator Consensus):**
  - **41.5%** were correctable **model errors**.
  - **58.5%** were **"inherently ambiguous"** cases (visual evidence is insufficient).

Error Analysis – Model vs. Ambiguity

- We manually analyzed all 785 errors from three representative folds (worst, median, and best).
- **Findings (Based on Annotator Consensus):**
  - **41.5%** were correctable **model errors**.
  - **58.5%** were **“inherently ambiguous”** cases (visual evidence is insufficient).
- **The Upper Bound:**
  - Based on this consensus, the estimated upper limit for the task is around **88.5% macro-accuracy**.
  - Our system (79.7%) closes a substantial part of the achievable gap.

Error Analysis – The Visual Limits of VCR

- Ambiguous errors occur due to low illumination, reflections, partial views, or borderline hues.

Pred: Black
GT: Brown

Pred: Gray
GT: Silver

Pred: Red
GT: Orange

Pred: Multicolored
GT: White

Pred: Multicolored
GT: White

Pred: White
GT: Silver

Pred: Multicolored
GT: Unknown

Pred: Brown
GT: Green

Conclusions

- **Synthetic data is highly effective** when paired with the right training objective (WCE) and domain alignment (image-conditioned with Gemini).

Conclusions

- **Synthetic data is highly effective** when paired with the right training objective (WCE) and domain alignment (image-conditioned with Gemini).
- By combining synthetic data, DINOv3 representations, WCE, LWCD, color-safe augmentation, and SAM 2 preprocessing, we improved state-of-the-art macro-accuracy by **8.2 percentage points**.

Conclusions

- **Synthetic data is highly effective** when paired with the right training objective (WCE) and domain alignment (image-conditioned with Gemini).
- By combining synthetic data, DINOv3 representations, WCE, LWCD, color-safe augmentation, and SAM 2 preprocessing, we improved state-of-the-art macro-accuracy by **8.2 percentage points**.
- **The practical limits of VCR** depend heavily on the amount and reliability of color evidence available in unconstrained surveillance imagery.

ICPR 2026
International Conference on Pattern Recognition
LYON

Thank you!

https://github.com/viniciusorru/vcr-synthetic

CAPES

ARAUCARIA
CIÊNCIA, TECNOLOGIA E INOVAÇÃO PARA O PARANÁ

PARANÁ
GOVERNO DO ESTADO
SECRETARIA DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

CNPq

PUCPR
GRUPO MARISTA

STJ SUPERIOR TRIBUNAL DE JUSTIÇA

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

NVIDIA