Toward Enhancing Vehicle Color Recognition in Adverse Conditions: A Dataset and Benchmark

Gabriel E. Lima*, Rayson Laroca[†,*], Eduardo Santos[‡,*], Eduil Nascimento Jr.[‡], and David Menotti*

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

*{gelima,menotti}@inf.ufpr.br [†]rayson@ppgia.pucpr.br [‡]{ed.santos,eduiljunior}@pm.pr.gov.br

Abstract—Vehicle information recognition is crucial in various practical domains, particularly in criminal investigations. Vehicle Color Recognition (VCR) has garnered significant research interest because color is a visually distinguishable attribute of vehicles and is less affected by partial occlusion and changes in viewpoint. Despite the success of existing methods for this task, the relatively low complexity of the datasets used in the literature has been largely overlooked. This research addresses this gap by compiling a new dataset representing a more challenging VCR scenario. The images – sourced from six license plate recognition datasets – are categorized into eleven colors, and their annotations were validated using official vehicle registration information. We evaluate the performance of four deep learning models on a widely adopted dataset and our proposed dataset to establish a benchmark. The results demonstrate that our dataset poses greater difficulty for the tested models and highlights scenarios that require further exploration in VCR. Remarkably, nighttime scenes account for a significant portion of the errors made by the best-performing model. This research provides a foundation for future studies on VCR, while also offering valuable insights for the field of fine-grained vehicle classification.

I. Introduction

Over the past two decades, there has been significant interest in extracting vehicle information from images taken by surveillance cameras. This technology plays a crucial role in various practical domains [1]–[4], especially in criminal investigations. Within this context, Vehicle Color Recognition (VCR) holds significant importance. Color covers a substantial portion of the vehicle's surface, making it less prone to partial occlusion and less affected by changes in viewpoint [5], [6].

Previous research on VCR can be broadly categorized into two main groups: handcrafted methods [1], [7]–[10] and deep learning-based approaches [2]–[6]. Alongside the approaches, several datasets have been created to tackle the VCR problem. Notably, Chen et al. [1] introduced the first publicly available dataset, comprising 15,601 vehicles categorized into eight color classes. This dataset has become a popular choice for subsequent studies [2]–[5] due to its original challenging nature, characterized by images with lighting variations, haze, and overexposure.

Despite these conditions, studies have reported satisfactory outcomes, achieving up to 95% average accuracy in color recognition across the mentioned dataset and datasets with similar characteristics [2]–[4], [6]. However, upon analyzing these studies, it becomes evident that the explored datasets

(a) Images from the dataset proposed by Chen et al. [1].

(b) Images from the UFPR-VCR dataset, proposed in this work.

Fig. 1. Examples of images from the datasets proposed in [1] (a) and in this work (b), with the corresponding vehicle color annotation shown above each image. Observe that images in the proposed dataset (b) depict significantly more challenging scenes than those in (a), featuring adverse conditions such as nighttime settings and vehicles from various viewpoints.

lack images depicting highly adverse conditions. The images primarily feature vehicles captured under adequate lighting conditions and from a consistent viewpoint, with clearly distinguishable colors, as shown in Fig. 1a. Consequently, the reported satisfactory results may be misleading, as the testing scenarios do not fully comprise the challenges often found in real-world, unconstrained VCR applications.

To tackle a more complex VCR scenario, we introduce the UFPR Vehicle Color Recognition (UFPR-VCR) dataset[1,2]. It comprises 10,039 images featuring various real-world conditions such as frontal and rear views, partially occluded

---

[1]https://github.com/lima001/ufpr-vcr-dataset
[2]Access is granted upon request, i.e., interested parties must register by filling out a registration form and agreeing to the dataset's terms of use.

vehicles, diverse and uneven lighting, and nighttime scenes. The images were sourced from six public datasets collected in Brazil, originally intended for Automatic License Plate Recognition (ALPR). The images cover eleven distinct vehicle colors, with annotations for over 90% of the vehicles validated using information obtained from the corresponding license plates (see details in Section III-C).

Fig. 1b presents sample vehicle images from the proposed dataset, along with their corresponding color annotations. These images highlight the challenging VCR scenarios considered in this work, demonstrating the difficulty of accurately classifying vehicle colors. To assess the dataset's challenges and pinpoint areas for improvement in VCR, we conduct a benchmark study using four deep learning-based models.

The remainder of this work is organized as follows. Section II reviews related works. Section III introduces the UFPR-VCR dataset. Section IV details the conducted experiments and presents the achieved results. Finally, Section V summarizes our findings and outlines directions for future research.

II. Related Work

This section overviews the datasets proposed within the VCR context. It outlines their key characteristics, describes the methodologies employed in studies using these datasets, and summarizes the results they have achieved.

In 2007, Baek et al. [7] proposed a dataset comprising 500 images from unspecified scenarios, equally distributed across five colors. They utilized a 2D histogram technique within the HSV color space for feature extraction and employed Support Vector Machines (SVMs) for classification, reporting an average accuracy of 94.9%. Using this dataset, Son et al. [8] further refined the SVM classifier by proposing a convolution kernel, achieving precision and recall rates exceeding 92%.

Three years later, Dule et al. [9] introduced a dataset containing 1,960 highway images, evenly distributed across seven colors. Half of the images were pre-processed to include a smoothed hood region of the vehicle, while the remaining images show frontal views. The highest accuracy (83.5%) was achieved using histogram features from different color spaces, classified with a multilayer perceptron.

While the datasets mentioned above were once accessible, they are currently unavailable to the best of our knowledge. In 2014, Chen et al. [1] introduced a dataset for VCR consisting of 15,601 frontal images captured by surveillance cameras under challenging conditions, compared to other datasets at the time. These images are categorized into eight colors: black, blue, cyan, gray, green, red, white, and yellow. The authors employed an implicit region-of-interest selector integrating spatial information for feature extraction. By combining this technique with principal component analysis and an SVM classifier, they attained an average recognition precision of 92.5%.

In subsequent studies using the same dataset, efforts were made to refine the feature extraction method. Hu et al. [5] integrated a Convolutional Neural Network (CNN) with Spatial Pyramid Pooling (SPP), achieving an average precision of 94.6%. Zhang et al. [2] further proposed feature fusion from multiple CNN layers via SPP, resulting in an average precision of 95.4%. Fu et al. [3] introduced the Multiscale Comprehensive Feature Fusion Convolutional Neural Network (MCFF-CNN), which incorporates residual learning and inception modules, achieving an average precision above 97%.

In 2021, Wang et al. [6] built a dataset featuring 32,220 vehicles distributed across eleven colors, further subdivided into 75 subcategories. The dataset exclusively contains rear-view vehicle images, annotated using a clustering algorithm and prior knowledge of typical vehicle colors. The researchers achieved an average accuracy of 97.8% using a hybrid model that combines CNN and Vision Transformer (ViT) models. Despite our efforts to contact the authors for access to the dataset, we have not yet received a response.

In 2023, Hu et al. [4] introduced the *Vehicle Color-24* dataset, composed of 31,232 vehicles categorized into 24 color classes. Before preprocessing, the dataset included 10,091 frontal-view vehicle images suitable for both vehicle detection and color identification tasks. The authors employed CNN and Feature Pyramid Networks (FPN) modules for multi-scale information fusion, alongside a loss function aimed at addressing the dataset's long-tail distribution. They reported a mean average precision of 95.0%.

Although initially appealing, the Vehicle Color-24 dataset was not considered in this research for two key reasons. First, we obtained access to the dataset only recently, during the course of this study. Second, all samples within the dataset underwent preprocessing steps, including haze removal and lighting adjustments, by its creators. These alterations may have reduced the dataset's diversity and, consequently, its ability to faithfully represent real-world conditions.

Despite the progress in VCR research with the development of increasingly robust methods, our analysis of existing datasets reveals that they predominantly feature relatively simple scenarios. These scenarios typically show vehicles in daylight hours, captured from a single perspective, without occlusions or vehicles partially outside the image frame. While these datasets have enabled methods to achieve high success rates, they do not fully capture the complexities of real-world scenarios. Therefore, the introduction of the UFPR-VCR dataset represents a distinctive contribution to advancing this field. Furthermore, previous research has focused on maximizing performance on the VCR task. This research, however, aims to identify the complex scenarios that limit VCR success in unconstrained scenarios and use those insights to drive further development in the field.

III. The UFPR-VCR Dataset

This section details the creation process and main characteristics of the UFPR-VCR dataset. It includes 10,039 images from 9,502 different vehicles, categorized into eleven colors: beige, black, blue, brown, gray, green, orange, red, silver, white, and yellow. The distribution of images across these colors is highly unbalanced (see Fig. 2). This reflects the real-world scenario, where vehicle colors such as white, black, and shades of gray are more common than others [11]–[13].
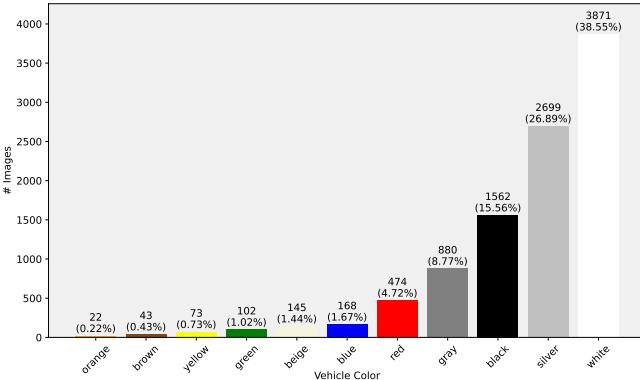
Fig. 2. Distribution of vehicle colors in the UFPR-VCR dataset.

The images show vehicles across different categories (i.e., cars, vans, buses, and trucks) captured in diverse environments. The original images, before vehicle cropping, were sourced from six Brazilian datasets commonly used in ALPR research, namely: OpenALPR-BR [14], RodoSol-ALPR [15], SSIG-SegPlate [16], UFOP [17], UFPR-ALPR [18], and Vehicle-Rear [19][3]. Table I provides an overview of these datasets.

TABLE I
SUMMARY OF THE ALPR DATASETS USED TO CREATE UFPR-VCR.

Dataset	Year	Images	Resolution	Viewpoint
UFOP [17]	2011	377	800×600	Frontal/Rear
SSIG-SigPlate [16]	2016	2,000	1920×1080	Frontal
OpenALPR-BR [14]	2016	115	Various	Frontal/Rear
UFPR-ALPR [18]	2018	4,500	1920×1080	Frontal/Rear
Vehicle-Rear* [19]	2021	445*	1280×720	Rear
RodoSol-ALPR [15]	2022	20,000	1280×720	Frontal/Rear

* We used only the portion of Vehicle-Rear that includes labels for the license plates.

We selected these datasets for three main reasons: (i) they are widely adopted [15], [20]; (ii) their images depict scenes with diverse lighting and viewpoints; and (iii) they include license plate-related annotations, enabling the validation of color annotations and therefore minimizing labeling errors.

As expected, the chosen datasets do not share a standard organization scheme. Hence, preprocessing and image selection procedures were implemented to standardize the images and identify those appropriate for the VCR task. These procedures are detailed in Section III-A and Section III-B, respectively. Lastly, Section III-C describes the labeling process.

A. Preprocessing

Vehicles were cropped and separated into individual images. Vehicle bounding box information was readily available in the OpenALPR-BR, SSIG-SegPlate, UFOP, UFPR-ALPR, and Vehicle-Rear datasets. Due to the lack of this information in the RodoSol-ALPR dataset, we used the well-known YOLOv8 model [21] for vehicle detection, as depicted in Fig. 3. This model was chosen for its robust performance and extensive use in both academic research and industry [21]–[23]. Importantly,

[3] We received permission from the creators of the explored ALPR datasets to build the UFPR-VCR dataset.

no noise removal or image enhancement techniques were applied to preserve the original adverse image conditions.

Fig. 3. Illustration depicting the process of extracting vehicle patches from images in the RodoSol-ALPR dataset, which lacks vehicle position labels.

B. Image selection

After preprocessing, a total of 28,061 images were obtained. However, another filtering process was necessary to ensure that every image in the UFPR-VCR dataset is suitable for the VCR task. This involved discarding 10,870 motorcycle images from UFPR-ALPR (870) and RodoSol-ALPR (10,000), as these images represent scenarios where identifying the vehicle color was nearly impossible (see three examples in Fig. 4).

| (a) Red | (b) Red | (c) Blue |

Fig. 4. Examples of discarded motorcycle images: (a) and (c) were sourced from the RodoSol-ALPR dataset [15], while (b) was extracted from the UFPR-ALPR dataset [18]. Below each image is the corresponding motorcycle's color. In this figure, the original images were slightly resized for better viewing.

Another selection process was conducted to remove redundant images from the tracks in the SSIG-SegPlate and UFPR-ALPR datasets. These datasets contain different vehicle tracks, each comprising a series of sequential frames extracted from videos focusing on a single target vehicle (although other vehicles may appear in the background) [16], [18]. Due to the low variability between vehicle images cropped from the same track, only one vehicle image per track was selected. The criterion for selection was to choose the middle image from each track (e.g., if a track contains 30 images, the 15th image was selected). As a result, 6,278 vehicle images were excluded, and 10,913 were retained.

A similar case occurs in the Vehicle-Rear dataset, as it also includes sequential frames extracted from videos. To address heavily occluded images resulting from overlapping vehicles within the camera's field of view, preference was given to images that clearly depict the vehicle's body. When multiple images met this criterion, selections were made randomly, leading to the exclusion of 373 images from the dataset. Following this process, 10,540 vehicle images were kept.

In the RodoSol-ALPR dataset, while not consisting of sequential frames extracted from videos, individual vehicles appear multiple times across different days and times. To ascertain whether images of the same vehicle represent distinct scenarios [24], images were grouped based on their license

plate annotations and manually verified. During this review, 418 images were excluded due to minimal perceptible differences in lighting, pose, or other distinguishing characteristics, while 10,122 images were retained.

The final selection process involved removing 83 images that lacked identifiable colors. This was mainly observed (i) when vehicles were heavily occluded or substantially outside the image frame; and (ii) when vehicles are registered as "multicolored"[4] (a rare occurrence). Fig. 5 provides examples of images excluded during this selection process.

(a) Multicolored   (b) White   (c) White

Fig. 5. Examples of images excluded due to vehicles with multiple colors (a) and those partially outside the image frame (b, c). The colors of the vehicles in (b) and (c) cannot be visually determined solely from their front bumpers. Note that the vehicle images in this figure were resized for improved visibility.

C. Annotations

As the UFPR-VCR dataset derives from established ALPR datasets, each image was initially associated with the license plate of the target vehicle. These annotations enabled the automated retrieval of vehicle information from Brazil's National Traffic Secretariat (SENATRAN) database, streamlining the annotation process. In total, 9,502 unique license plates were identified. However, there were 212 plates (corresponding to 906 images) for which vehicle information was unavailable. These cases required manual annotation. Finally, vehicles with similar colors were grouped, and the annotations were manually re-validated to ensure the accuracy of each vehicle's label.

IV. EXPERIMENTS

This section presents a benchmark study conducted on the UFPR-VCR dataset. The study compares four deep learning models on the proposed dataset to evaluate its complexity and identify potential areas for improving VCR. Given the absence of prior studies demonstrating the suitability of these models for VCR, a brief study was performed on the dataset proposed by Chen et al. [1], the results of which are also presented here.

The rest of this section is structured as follows. Section IV-A covers the experimental methodology. Section IV-B presents the results obtained on the dataset proposed by Chen et al. [1]. Lastly, Section IV-C presents the results achieved on UFPR-VCR, highlighting the challenges posed by this dataset.

A. Methodology

The evaluation on the Chen et al. [1] dataset and the benchmark on the UFPR-VCR dataset share nearly identical methodology. The only difference is the training protocols applied. The materials and methods are summarized as follow:

---

[4]The term "multicolored" (a non-literal translation of '*fantasia*' in Portuguese) is used when the vehicle's primary color cannot be determined [25].

1) Models: we explored EfficientNet-V2 [26], MobileNet-V3 [27], ResNet-34 [28], and ViT b16 [29]. These models were chosen for their widespread adoption in computer vision tasks and their availability with implementations across various frameworks, which enhances research reproducibility.

2) Dataset splitting: the datasets were divided into training, validation, and test subsets using an 8:1:1 ratio. The images were distributed across the different colors for each subset, aiming to mirror the original dataset's class distribution as closely as possible. When the class distribution did not evenly divide between validation and test subsets, any surplus images were randomly assigned to one of the subsets.

3) Preprocessing: all images were resized to 224×224 pixels to align with the input size required by the models. To increase data variability during training, the following transformations were applied to each image in every training batch:

- Affine transformations (with a probability $p = 50\%$), including rotations ($\pm 180°$), scaling (from 0.9 to 1.3), and shearing ($\pm 180°$);
- Random adjustments to image brightness and contrast ($p = 30\%$), within a limit of 0.2;
- Blur using a generalized normal filter with randomly selected parameters ($p = 40\%$);
- A random 72×72 pixel section of the image is replaced with random noise ($p = 25\%$).

Finally, every image was normalized using the mean and standard deviation from ImageNet [30].

4) Training: We employed transfer learning by initializing all models with pre-trained weights from ImageNet [30]. Each network's final fully connected layer was adapted to produce outputs specific to the classes within the dataset being explored. To be precise, weight adjustments were confined to these layers during the training phase.

The Adam optimizer [31] was employed with $\beta_1 = 0.9$, $\beta_2 = 0.999$, a batch size of 128, weight decay set to 10^{-5}, and an initial learning rate of 10^{-4}. A learning rate reduction scheme was used with a patience value of 10 and a reduction factor of 10^{-1} upon plateau detection. Training extended up to a maximum of 400 epochs, with early stopping configured to halt training if no improvement was observed for 15 consecutive epochs. We used the cross-entropy loss function.

Two different training protocols were implemented for the UFPR-VCR dataset: (i) training with data augmentation only; and (ii) training with oversampling of minority classes to balance the dataset distribution. The adopted oversampling method increases the frequency of minority class samples during training by creating synthetic data through data augmentation. This technique is known to enhance results on imbalanced datasets [32], [33]. For the dataset proposed by Chen et al. [1], only protocol (i) was adopted, as it proved sufficient to achieve good results.

5) Evaluation: Each experiment was repeated five times using different dataset splits, and the results are reported based on the average outcomes. The evaluation metrics used are top-1 accuracy (Top-1), top-2 accuracy (Top-2), precision, recall, and F1-score (F1). These metrics were calculated globally for

each iteration using macro averaging. Running the experiments multiple times helps ensure the reliability of the results, reducing the influence of random variations in the data splits.

B. Results on the dataset proposed by Chen et al. [1]

This section aims to "pre-validate" the models adopted as benchmarks for UFPR-VCR, demonstrating their suitability for the VCR task. Table II presents the results obtained for each model on the dataset presented in [1]. Notably, the best-performing model is ViT b16, achieving an average precision close to those reported in the literature [1], [3], [5]. Furthermore, the Top-1 and Top-2 accuracies from the evaluated models indicate promising results for the VCR problem.

TABLE II
GLOBAL METRICS (%) REACHED BY ALL MODELS ON THE DATASET PROPOSED BY CHEN ET AL. [1] (AVERAGED OVER FIVE RUNS). THE MODELS WERE TRAINED WITH THE DATA AUGMENTATION PROTOCOL.

Model	Top-1	Top-2	Precision	Recall	F1
EfficientNet-V2 [26]	84.6	93.4	84.5	84.6	84.4
MobileNet-V3 [27]	90.6	96.7	91.7	90.6	91.0
ResNet-34 [28]	89.0	95.6	91.1	89.0	89.9
ViT b16 [29]	**92.8**	**98.0**	**95.3**	**92.8**	**93.9**

The results indicate that the explored models are suitable for studying the VCR problem, reinforcing the relevance of the proposed research. Specifically, with relatively minimal effort, we achieved results comparable to state-of-the-art works on the dataset introduced in [1]. Hence, we claim that research using this dataset (and others with similar characteristics) does not represent a challenging scenario for VCR evaluation.

C. Results on the UFPR-VCR dataset

Table III presents the results for each model on the UFPR-VCR dataset, considering the two training protocols. Remarkably, models trained using protocol (i) achieved better precision and F1 values but exhibited lower accuracy compared to those trained under protocol (ii). This discrepancy stems from the higher frequency of minority classes in protocol (ii). While this improves overall accuracy, it reduces the precision for the majority classes. As the test set is also unbalanced, the precision on models trained with protocol (ii) is negatively affected.

TABLE III
GLOBAL METRICS (%) REACHED BY ALL MODELS ON THE UFPR-VCR DATASET (AVERAGED OVER FIVE RUNS). PROTOCOL (II) INCORPORATES OVERSAMPLING OF MINORITY CLASSES, WHEREAS (I) DOES NOT.

Protocol	Model	Top-1	Top-2	Precision	Recall	F1
(i)	EfficientNet-V2 [26]	51.2	65.3	65.2	51.2	53.5
	MobileNet-V3 [27]	50.5	65.4	65.8	50.5	53.1
	ResNet-34 [28]	49.1	60.3	64.3	49.1	52.4
	ViT b16 [29]	**59.2**	**71.3**	**76.0**	**59.2**	**62.8**
(ii)	EfficientNet-V2 [26]	55.4	69.5	43.5	55.4	44.6
	MobileNet-V3 [27]	59.3	73.3	42.6	59.4	45.2
	ResNet-34 [28]	59.3	72.9	47.8	59.3	49.9
	ViT b16 [29]	**66.2**	**79.7**	**55.7**	**66.2**	**57.8**

Keeping this in mind, we analyzed correct and incorrect predictions using the best model in terms of top-1 accuracy, specifically ViT b16 trained under protocol (ii). Fig. 6 shows the normalized confusion matrix averaged across the five runs. It reveals that colors such as yellow, white and red were consistently identified with high accuracy compared to other classes. Conversely, colors such as brown, blue, green and gray posed challenges for identification. This difficulty likely stems from dataset characteristics, including significant variations in tone and lighting conditions, potentially causing these colors to be mistaken for shades of gray or black.
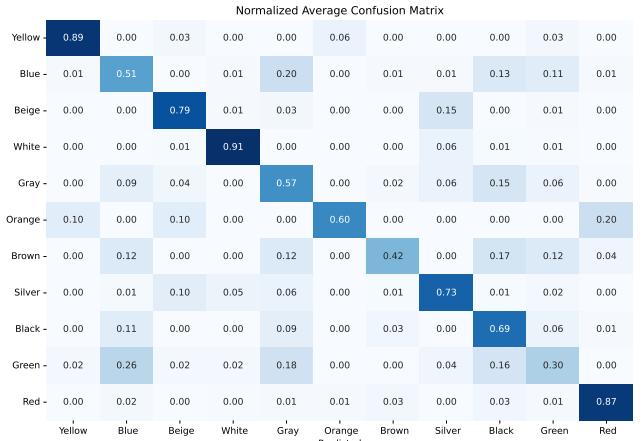
Fig. 6. Normalized confusion matrix illustrating the performance of the ViT b16 model trained with data augmentation and oversampling techniques.

Our analysis uncovered an interesting trend: nighttime images were misclassified at a much higher rate than daytime images. Specifically, 72 out of 222 misclassified images (32.4%) were captured at night. Nighttime images likely constitute less than 10% of the UFPR-VCR dataset, although the exact percentage is unknown as capture times are not labeled in the original datasets. This discrepancy is likely due to the inherent challenges of nighttime scenes, such as high illumination and overexposure from vehicle headlights (see Fig. 7). The causes behind the remaining misclassifications were less apparent.

GT: White   **GT:** Red   **GT:** Black
Pred: Silver   **Pred:** White   **Pred:** Gray

Fig. 7. Examples of nighttime images that were misclassified.

Finally, we analyzed the model's second-choice predictions (top 2) for images it initially misclassified (errors in terms of top-1 accuracy). Notably, colors like beige, white, gray, silver, and black achieved over 50% classification accuracy in these cases. In other words, the model's second prediction was correct for more than half of the misclassified images in these color categories. However, it is important to note that an average of 44.4% of the nighttime images remained incorrectly classified even considering the top-2 predictions.

V. Conclusions

This study revealed shortcomings in existing Vehicle Color Recognition (VCR) datasets, emphasizing their inadequacy in replicating real-world, unconstrained scenarios. To address this issue, we compiled the UFPR-VCR dataset. It comprises 10,039 images featuring adverse scenarios, such as various viewpoints, uneven lighting, and nighttime scenes, across 11 color classes. A benchmark study using four deep learning models demonstrated that UFPR-VCR presents significant challenges for VCR, particularly in nighttime scenarios, which accounted for $\approx$ 33% of the errors by the best-performing model despite representing a much smaller portion of the dataset.

This study identifies remaining scenarios in VCR that require further investigation. Developing novel methods for robust color recognition under adverse conditions is essential to improve the reliability of these approaches in real-world applications. We hope this work serves as a catalyst for VCR research in adverse conditions, encouraging future studies to address progressively more difficult scenarios.

An important future research direction is tackling the challenges of VCR in nighttime scenes. This endeavor would likely involve investigating advanced preprocessing methods and designing specialized architectures to enhance current results. Additionally, we plan to enrich the dataset by incorporating more vehicle attributes, such as type (e.g., sedan, hatchback, truck), make, and model. This would enable the integration of color recognition with fine-grained vehicle classification tasks, potentially through a multi-task learning framework.

References

[1] P. Chen, X. Bai, and W. Liu, "Vehicle color recognition on urban road by feature context," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, no. 5, pp. 2340–2346, 2014.

[2] Q. Zhang *et al.*, "Vehicle color recognition using multiple-layer feature representations of lightweight convolutional neural network," *Signal Processing*, vol. 147, pp. 146–153, 2018.

[3] H. Fu *et al.*, "MCFF-CNN: Multiscale comprehensive feature fusion convolutional neural network for vehicle color recognition based on residual learning," *Neurocomputing*, vol. 395, pp. 178–187, 2020.

[4] M. Hu, L. Bai, J. Fan, S. Zhao, and E. Chen, "Vehicle color recognition based on smooth modulation neural network with multi-scale feature fusion," *Frontiers of Computer Science*, vol. 17, no. 3, p. 173321, 2023.

[5] C. Hu, X. Bai, L. Qi, P. Chen, G. Xue, and L. Mei, "Vehicle color recognition with spatial pyramid deep learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2925–2934, 2015.

[6] Y. Wang *et al.*, "Transformer based neural network for fine-grained classification of vehicle color," in *International Conference on Multimedia Information Processing and Retrieval (MIPR)*, 2021, pp. 118–124.

[7] N. Baek, S.-M. Park, K.-J. Kim, and S.-B. Park, "Vehicle color classification based on the support vector machine method," in *International Conference on Intelligent Computing*, 2007, pp. 1133–1139.

[8] J.-W. Son, S.-B. Park, and K.-J. Kim, "A convolution kernel method for color recognition," in *International Conference on Advanced Language Processing and Web Information Technology*, 2007, pp. 242–247.

[9] E. Dule *et al.*, "A convenient feature vector construction for vehicle color recognition," in *WSEAS International Conference on Neural Networks, Evolutionary Computing and Fuzzy systems*, 2010, p. 250–255.

[10] J.-W. Hsieh *et al.*, "Vehicle color classification under different lighting conditions through color correction," *IEEE Sensors Journal*, vol. 15, no. 2, pp. 971–983, 2015.

[11] Axalta Coating Systems, "Global Automotive 2022 Color Popularity Report," https://www.axalta.com/refinisheurope_eu/en_GB/about-axalta/colour-popularity-report-axalta.html, 2022, accessed: 2024-06-28.

[12] V. Farias and G. Croquer, "Por que o carro colorido sumiu? 67% dos veículos no Brasil são brancos, pretos ou cinzas," https://g1.globo.com/economia/noticia/2023/08/20/por-que-o-carro-colorido-sumiu-67percent-dos-veiculos-no-brasil-sao-brancos-pretos-ou-cinzas.ghtml, 2023, accessed: 2024-06-28.

[13] M. Harley, "Estudo revela as cores mais populares e quais aumentam valor de revenda," https://forbes.com.br/forbeslife/forbes-motors/2023/10/estudo-revela-as-cores-mais-populares-e-quais-aumentam-valor-de-revenda/, 2023, accessed: 2024-06-28.

[14] OpenALPR, "OpenALPR-BR dataset," https://github.com/openalpr/benchmarks/tree/master/endtoend/br/, 2016.

[15] R. Laroca *et al.*, "On the cross-dataset generalization in license plate recognition," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, Feb 2022, pp. 166–178.

[16] G. R. Gonçalves, S. P. G. da Silva, D. Menotti, and W. R. Schwartz, "Benchmark for license plate character segmentation," *Journal of Electronic Imaging*, vol. 25, no. 5, p. 053034, 2016.

[17] P. R. Mendes Júnior *et al.*, "Towards an automatic vehicle access control system: License plate location," in *IEEE International Conference on Systems, Man, and Cybernetics*, Oct 2011, pp. 2916–2921.

[18] R. Laroca *et al.*, "A robust real-time automatic license plate recognition based on the YOLO detector," in *International Joint Conference on Neural Networks (IJCNN)*, July 2018, pp. 1–10.

[19] I. O. Oliveira *et al.*, "Vehicle-Rear: A new dataset to explore feature fusion for vehicle identification using convolutional neural networks," *IEEE Access*, vol. 9, pp. 101 065–101 077, 2021.

[20] R. Laroca *et al.*, "Leveraging model fusion for improved license plate recognition," in *Iberoamerican Congress on Pattern Recognition (CIARP)*, Nov 2023, pp. 60–75.

[21] Ultralytics, "YOLOv8," 2023, accessed: 2024-06-28. [Online]. Available: https://github.com/ultralytics/ultralytics

[22] J.-H. Kim, N. Kim, and C. S. Won, "High-speed drone detection based on YOLOv8," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023, pp. 1–5.

[23] F. M. Talaat and H. ZainEldin, "An improved fire detection approach based on YOLOv8 for smart cities," *Neural Computing and Applications*, vol. 35, no. 28, pp. 20 939–20 954, 2023.

[24] R. Laroca *et al.*, "Do we train on test data? The impact of near-duplicates on license plate recognition," in *International Joint Conference on Neural Networks (IJCNN)*, June 2023, pp. 1–8.

[25] Brasil, "Resolução nº 292, de 29 de Agosto de 2008," https://www.gov.br/infraestrutura/pt-br/assuntos/transito/conteudo-contran/resolucoes/cons292.pdf, 2008, accessed: 2024-06-28.

[26] M. Tan and Q. Le, "EfficientNetV2: Smaller models and faster training," in *International Conf. on Machine Learning*, 2021, pp. 10 096–10 106.

[27] A. Howard *et al.*, "Searching for MobileNetV3," in *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 1314–1324.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778.

[29] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021, pp. 1–12.

[30] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 248–255.

[31] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[32] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique," *Journal of Artificial Intelligence Research*, vol. 16, pp. 321–357, 2002.

[33] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, p. 60, 2019.