A Study on Fine-Grained Motorcycle Classification for Intelligent Transportation Systems

Gabriel E. Lima*, Eduardo Santos†,*, Eduil Nascimento Jr.†, Rayson Laroca‡,* and 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
*{gelima,menotti}@inf.ufpr.br †{ed.santos,eduiljunior}@pm.pr.gov.br ‡rayson@ppgia.pucpr.br

***Abstract*—Vehicle recognition from images is crucial to Intelligent Transportation System (ITS), supporting applications such as tolling, access control, and forensics. Fine-Grained Vehicle Classification (FGVC) extends this capability by identifying vehicles by make, model, and type. However, research has largely centered on four-wheeled vehicles, with motorcycles receiving limited attention despite representing a substantial share of traffic in many countries, including Brazil. This under-representation can reduce ITS effectiveness and fairness. This work-in-progress study addresses this gap by investigating Fine-Grained Motorcycle Classification (FGMC) in real-world ITS scenarios. We evaluate seven deep learning architectures under two training protocols for independent make and model classification. To enable this, we augment a widely adopted dataset for Automatic License Plate Recognition (ALPR) with motorcycle make and model annotations. Results show that FGMC is feasible within the studied context, yet performance is hindered by severe class imbalance, underscoring the need for improved balancing strategies. The results also reveal a drop in accuracy under challenging conditions, particularly at night or in low-light environments. Future directions include expanding the dataset to more diverse scenarios and exploring FGMC integration with ALPR to enhance overall vehicle identification accuracy.**

I. Introduction

Recognizing vehicles from images is a key task in Intelligent Transportation System (ITS), enabling applications such as traffic management, automatic toll collection, and surveillance. Fine-Grained Vehicle Classification (FGVC) focuses on identifying vehicle attributes — such as make, model, and type — which can enhance ITS solutions directly [1], [2] or complement technologies like Automatic License Plate Recognition (ALPR) to improve vehicle identification [3], [4].

Over the past decade, FGVC has advanced through deep learning and the introduction of fine-grained vehicle datasets [5], [6]. Nonetheless, research has focused predominantly on four-wheeled vehicles, with motorcycles receiving little attention. When considered, motorcycles are usually classified into broad categories such as vehicle type (car, motorcycle, truck) or motorcycle class (motorcycle, scooter, tricycle) [7]–[9].

The few works addressing finer-grained labels often rely on web-sourced images or do not assess motorcycles independently from other vehicle types [10], [11]. While these efforts mark an initial step toward Fine-Grained Motorcycle Classification (FGMC), the task and its challenges remain insufficiently understood.

This oversight is concerning, as motorcycles represent a substantial share of the vehicle fleet in many countries. In Brazil, for example, the motorcycle fleet reached approximately 35 million vehicles, reflecting a 42% growth between 2014 and 2024 [12]. At the same time, thefts targeting specific motorcycle models have become more frequent [13]. Inadequate representation of motorcycles in ITS applications can compromise system accuracy and fairness, especially when license plates are occluded or illegible [14]. Reliable recognition of motorcycle make and model can strengthen vehicle identification and support verification against official records, which is essential for forensic and security purposes.

In response, this work-in-progress study aims to investigate FGMC in depth. Specifically, we evaluate seven deep learning architectures using two distinct training protocols to independently classify motorcycle make and model. To support this task, we extend a widely adopted ALPR dataset by incorporating detailed motorcycle annotations, ensuring alignment with ITS scenarios. Beyond reporting classification performance, we perform an error analysis to better understand both the overall behavior of the models and the challenges inherent to by FGMC.

The results highlight that FGMC is more challenging than traditional vehicle classification. Motorcycles present specific difficulties, such as limited availability of rear-view images in surveillance and fewer distinctive visual features. We hope this study provides a foundation for future research, encouraging the development of dedicated methods and datasets to improve motorcycle recognition in real-world applications.

The remainder of this paper is structured as follows. Section II reviews related work. Section III outlines the data preparation process. Section IV describes the experimental setup, presents the results, and discusses the initial limitations and challenges of FGMC. Finally, Section V concludes the study and outlines directions for future research.

II. Related Work

FGVC has received substantial attention over the past decade, with most research centered on four-wheeled vehicles. This section first reviews key developments in the field, highlighting the breadth and depth of existing work. We then shift focus to the relatively few studies on motorcycle classification and position our work within this context.

Accepted for presentation at SIBGRAPI 2025 – Workshop of Works in Progress (WiP).

Early FGVC research relied on hand-crafted features combined with traditional machine learning classifiers [15], [16]. With the rise of deep learning and the availability of large public datasets [5], [6], Convolutional Neural Networks (CNNs) became the dominant approach, recently complemented by Vision Transformers (ViTs). Much of the literature has focused on enhancing classification accuracy and tackling broader challenges, such as out-of-distribution recognition within FGVC.

Despite these advances, fine-grained classification of motorcycles remains largely unexplored. In most vehicle classification research that includes motorcycles, they are handled as part of coarse vehicle-type recognition tasks, distinguishing broad categories like cars, buses, and motorcycles. Only a few studies, notably Khoba et al. [10] and Roomi et al. [11], have investigated fine-grained motorcycle recognition focusing on make and model classification.

In 2023, Khoba et al. [10] introduced the Fine-Grained Vehicle Detection dataset, comprising $5,502$ images annotated with 210 fine-grained classes and bounding boxes for $24{,}450$ vehicles. Although motorcycles were not the primary focus, the dataset includes two motorcycle makes and 11 models. Their study evaluated object detection models and a hierarchical YOLOv5 variant. In 2024, Roomi et al [11] compiled a dataset of $5{,}000$ motorcycle images from various sources covering 27 motorcycle classes. They also approached the task using object detection models.

While these studies represent progress toward FGMC, important gaps remain. The dataset explored in [11] lacks the realistic conditions typical of ITS, and the absence of error analysis limits insight into the intrinsic challenges of motorcycle make/model classification. The research described in [10] is more comprehensive and includes error analysis, but it considers relatively few motorcycle classes and does not examine them in depth. Moreover, neither study evaluates standard deep learning classification architectures commonly adopted in FGVC.

This work-in-progress study addresses these gaps by investigating FGMC in real-world ITS scenarios. We evaluate seven deep learning architectures and analyze the broader characteristics and patterns of motorcycle recognition present in this context. We aim to establish a foundation for future research, fostering the development of specialized methods and datasets to advance FGMC in practical applications.

III. Data Preparation

This section describes the data process for adapting an ALPR dataset (originally intended for license plate detection and recognition) to the FGMC task. We selected the RodoSol-ALPR dataset [17], which contains $20{,}000$ images captured by static cameras at toll booths on a Brazilian highway (see Fig. 1). Of these, $10{,}000$ images depict rear-view motorcycles. The dataset was chosen based on: (i) its widespread use in vehicle identification research [18]–[21]; (ii) its strong alignment with a real-world ITS scenario; and (iii) its accurate license plate annotations, which facilitate FGMC labeling and enable future integration with ALPR research.

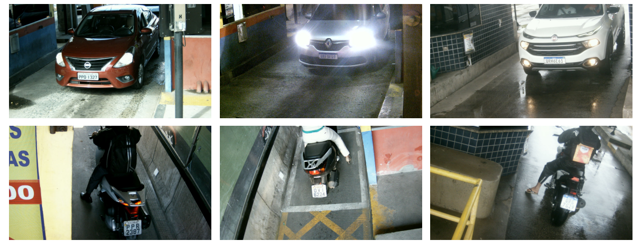


Fig. 1. Samples from the RodoSol-ALPR dataset [17], showcasing the diversity of vehicles and lighting conditions. Images have been slightly resized for improved visualization.

To prepare the dataset for FGMC, we applied filtering, standardization, and annotation steps. First, we removed all non-motorcycle images, retaining the $10{,}000$ original motorcycle samples. Next, we excluded images where the motorcycle was not clearly visible; that is, images captured under extremely low light, with heavy occlusion, or with the motorcycle largely out of frame (see Fig. 2). Tricycles were also excluded, as they accounted for less than 0.003% of the data, and the focus is on two-wheeled vehicles. After filtering, $8{,}556$ images remained.

(a) (b) (c)

Fig. 2. Motorcycle samples excluded due to: (a) extreme low-light conditions; (b) heavy occlusion caused by its cargo box; and (c) motorcycle being largely out of frame. Images have been slightly resized for better viewing.

To ensure consistency, all images were standardized to show only the motorcycle, excluding background elements. As the dataset lacked bounding box annotations, we used the YOLOv11-small model [22] to detect and crop the motorcycles. This model was chosen due to YOLO's strong performance, wide adoption in both academia and industry [23]–[25], and its pretrained support for the motorcycle class. In 358 images where the model failed to produce valid detections, the motorcycles were manually cropped.

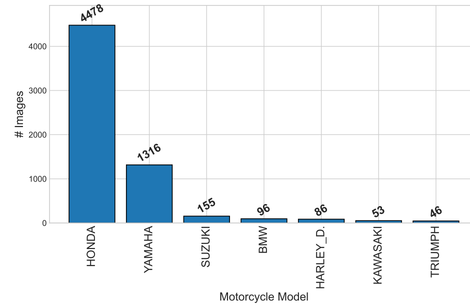
To reduce redundancy, motorcycle images were grouped by license plate and sorted by filename. A pruning heuristic was then applied based on the number of images I available for each motorcycle. If $I \leq 3$, all images were retained. If $3 < I < 10$, only the first and last images were kept. If $I \geq 10$, the first and last images were kept along with 10% of the remaining images, evenly spaced across the set. Each motorcycle was limited to a maximum of five images, with manual filtering applied when the automatic pruning exceeded this limit. This yielded a representative subset of $6{,}511$ images.

We then carried out the annotation process by linking each image to its corresponding license plate, thereby enabling the automatic retrieval of vehicle metadata from a record database via an API service. Out of $5{,}167$ unique license plates, 35 lacked complete information and were manually annotated. With all images labeled, we proceeded to generate two datasets — one for motorcycle make recognition and another for model recognition — as the latter required slightly different refinement criteria.

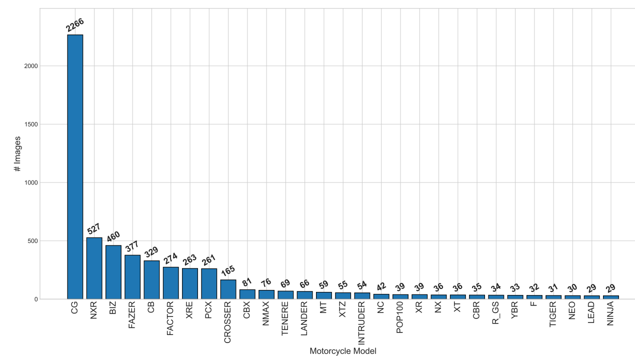
For model recognition, we adopted a higher-level grouping strategy by categorizing motorcycles at the model-family level rather than distinguishing between submodels. For example, "*Honda CG-125*" and "*Honda CG-150*" were grouped under a single class. This approach reduced class imbalance and avoided underrepresented categories. However, it also required the exclusion of Harley-Davidson motorcycles, whose models could not be grouped consistently by us. Lastly, we removed any make or model class with fewer than 25 unique vehicles.

The final output consists of two FGMC datasets[1]: MotorcycleMake, with 6,230 images across 7 make classes; and MotorcycleModel, with 5,827 images across 29 model classes. Fig. 3 shows sample images shared by both datasets, illustrating the variety of conditions and the challenge of distinguishing visually similar motorcycles from a limited viewing perspective. Finally, Fig. 4 presents the class distribution of each dataset, revealing a long-tailed pattern typical of Brazilian traffic, where a few makes/models dominate [26].

Fig. 3. Samples appearing in both the *MotorcycleMake* and *MotorcycleModel* datasets. The first and second text rows below each image show the motorcycle make and model, respectively. These examples illustrate variations in lighting conditions and capture angles. Images have been slightly resized for clarity.



(a) Class distribution in the *MotorcycleMake* dataset.



(b) Class distribution in the *MotorcycleModel* dataset.

Fig. 4. Class distributions for the proposed datasets, highlighting their long-tailed nature, where a small number of classes dominate.

IV. Experiments and Results

This section presents the experimental evaluation of deep learning models for motorcycle make and model recognition. The objective is to assess model performance on the *MotorcycleMake* and *MotorcycleModel* datasets separately and to examine the particularities of each task. The section is organized as follows: Section IV-A outlines the experimental methodology, and Section IV-B reports the classification results and provides an analysis based on the best-performing classifier.

A. Methodology

For each classification task, we evaluated seven deep learning architectures: EfficientNet-V2 [27], MobileNet-V3 [28], ResNet-50 [29], ResNet-101 [29], SwinTransformer-V2 [30], ViT-B16 [31], and YOLOv11-nano-cls [32] (a variant of YOLO for classification). These models were chosen due to their proven effectiveness in computer vision, widespread adoption [19], [25], and availability of open-source implementations to facilitate reproducibility.

All models were initialized with ImageNet-pretrained weights and fine-tuned with all layers unfrozen. The final fully connected layer was replaced to match the number of classes in the target dataset, and training was performed using the cross-entropy loss function. Training was allowed to run for up to 1,000 epochs, with early stopping activated if no improvement was observed on the validation set for 100 consecutive epochs.

For the standard models, the Adam optimizer was employed with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$, a weight decay of 5×10^{-4}, an initial learning rate of 10^{-2}, a minimum learning rate of 10^{-8}, and a batch size of 64. The learning rate was reduced by a factor of 10 if the validation loss plateaued for 25 epochs. The YOLOv11 models followed the Ultralytics training protocol, utilizing stochastic gradient descent with three warm-up epochs and a cyclic learning rate oscillating between 10^{-2} and 10^{-4} following the cosine decay.

[1]The datasets used in this study, derived from RodoSol-ALPR [17] and comprising a subset of images with FGMC annotations and corresponding splits, are available upon request, with instructions for obtaining them provided at: `https://github.com/Lima001/UFPR-FGMC-Dataset`.

TABLE I

Classification performance (%) of all models on the *MotorcycleMake* and *MotorcycleModel* datasets. Results show the average of 10 runs with standard deviations in parentheses; best outcomes are highlighted in bold. Protocol (p2) employs balanced sampling, while protocol (p1) follows standard training without balancing.

Models trained with protocol (p1)

Model	*MotorcycleMake*			*MotorcycleModel*		
	mi-acc	ma-acc	F1	mi-acc	ma-acc	F1
EfficientNet-V2 [27]	**95,37 (0,59)**	**77,09 (3,51)**	**80,35 (2,29)**	**94,62 (0,59)**	**85,50 (2,13)**	**86,83 (1,74)**
MobileNet-V3 [28]	93,49 (0,50)	73,69 (2,59)	77,53 (2,36)	91,88 (0,74)	78,69 (3,05)	80,67 (2,59)
ResNet-50 [29]	90,50 (1,24)	57,61 (6,71)	59,31 (7,69)	92,80 (0,73)	81,16 (2,71)	82,39 (2,42)
ResNet-101 [29]	89,48 (0,64)	54,08 (2,23)	55,71 (3,10)	92,18 (1,17)	80,00 (3,18)	81,60 (3,34)
SwinTransformer-V2 [30]	76,92 (0,98)	37,35 (3,45)	41,99 (4,42)	67,73 (1,85)	36,22 (2,11)	40,47 (2,34)
ViT-B16 [31]	79,15 (0,82)	46,10 (3,67)	52,08 (3,64)	70,60 (1,06)	44,07 (1,53)	48,60 (2,00)
YOLOv11-nano-cls [32]	93,25 (0,58)	71,91 (2,99)	75,88 (2,86)	91,18 (1.00)	77,54 (3,86)	78,90 (2,90)

Models trained with protocol (p2)

Model	*MotorcycleMake*			*MotorcycleModel*		
	mi-acc	ma-acc	F1	mi-acc	ma-acc	F1
EfficientNet-V2 [27]	**94,65 (0,62)**	**77,55 (4,14)**	**81,19 (3,64)**	**93,29 (0,60)**	**81,21 (2,24)**	**83,97 (2,02)**
MobileNet-V3 [28]	91,42 (0,47)	70,11 (3,36)	74,72 (2,52)	90,65 (0,49)	76,08 (2,59)	79,14 (1,96)
ResNet-50 [29]	92,05 (1,11)	71,95 (3,43)	75,52 (2,96)	91,64 (0,73)	78,58 (3,18)	81,30 (2,45)
ResNet-101 [29]	91,16 (1,19)	70,62 (1,88)	74,07 (1,87)	90,61 (1,13)	76,30 (2,64)	78,67 (2,10)
SwinTransformer-V2 [30]	70,39 (3,10)	14,22 (0,24)	12,16 (0,73)	24,74 (13,8)	03,54 (0,43)	01,56 (0,56)
ViT-B16 [31]	70,50 (2,70)	14,25 (0,06)	11,91 (0,08)	23,47 (13,4)	17,10 (13,5)	12,96 (12,0)
YOLOv11-nano-cls [32]	91,87 (0,82)	71,38 (3,93)	73,36 (3,25)	89,39 (1,44)	76,25 (2,70)	77,48 (2,73)

All models were trained using the default YOLO image classification augmentation pipeline [33], which consists of random resized cropping to 224×224 pixels, horizontal flipping, and RandAugment [34]. The latter applies two randomly selected augmentations from a fixed set of geometric and photometric transformations at a predetermined magnitude. During inference, input images are resized while preserving their aspect ratio, then center-cropped to 224×224 pixels and normalized.

We adopted two training strategies: (p1) standard training with data augmentation, and (p2) training with data augmentation and balanced sampling. In the second protocol, each class contributes exactly 300 samples per epoch. Classes with fewer images were oversampled by repeating samples before augmentation, while larger classes were downsampled randomly. This method maintains a balanced class distribution during training, aiming to mitigate bias toward majority classes.

The dataset was split into five non-overlapping folds, ensuring that all images of the same motorcycle (identified by its license plate) remained within the same fold to avoid data leakage [35]. The original class distribution was approximately maintained across folds. From these, 10 unique train-validation-test splits were generated by combining folds in a 3:1:1 ratio, such that each fold served as the test set twice. This approach ensured a thorough and balanced evaluation across the entire dataset.

To evaluate classification performance, we report macro accuracy (ma-acc), micro accuracy (mi-acc), and F1-score (F1). These metrics capture both per-class and overall performance, providing a balanced assessment across different class distributions. Final results are averaged over the dataset splits and independently reported for *MotorcycleMake* and *MotorcycleModel*. Since results can be close in value, we apply the Wilcoxon signed-rank test [36] ($p < 0.05$) to compare the models. This non-parametric test was chosen because it does not assume normality, is suitable for paired data, and is robust for small sample sizes [37].

B. Results

Table I reports classification results for all models on both motorcycle make and motorcycle model recognition, under training protocols (p1) and (p2). The highest score for each metric in each scenario is shown in bold. EfficientNet-V2 achieved the best overall performance across all metrics and scenarios. In contrast, attention-based models (i.e., ViT and SwinTransformer) performed the worst, likely due to their dependence on large training datasets and the class imbalance in both datasets.

Balanced sampling in protocol (p2) produced mixed effects, depending on the model architecture. For EfficientNet-V2, MobileNet-V3, and YOLO11-nano-cls, performance changes were generally small and slightly negative. Attention-based models experienced notable drops in all metrics, reinforcing their sensitivity to limited and imbalanced data. Conversely, both ResNet-based models benefited in the make recognition task, with macro metrics improving from below 60% to above 70%.

Model recognition consistently outperformed make recognition, even though the latter is a coarser classification task. A likely explanation is the high intra-class variability in makes, as different models from the same make can look substantially different. This issue is amplified in the *MotorcycleMake* dataset, where models from certain makes share visual characteristics with models from other makes.

Next, we analyze motorcycle behavior in the proposed ITS scenario by focusing on the best-performing model, EfficientNet-V2 trained under protocol (p1), which achieved the highest overall metrics. First, we examine common classification errors for each task, then identify patterns and factors influencing performance.

Fig. 5 shows the averaged confusion matrix for motorcycle make recognition. As expected, the two most frequent classes — Honda and Yamaha — achieved the highest accuracy. BMW and Harley-Davidson ranked next, performing better than Suzuki, Triumph, and Kawasaki, which showed noticeably lower accuracy. This is likely because BMW and Harley-Davidson have distinctive and consistent design languages across models, unlike the lower-performing makes.

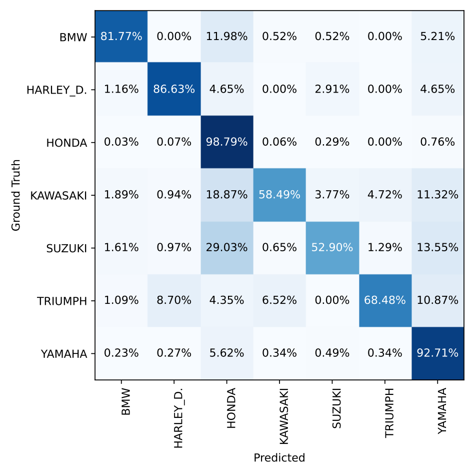
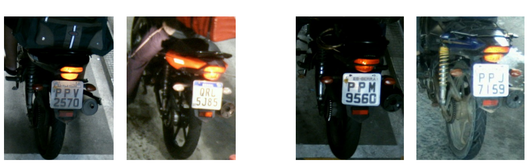


Fig. 5. Normalized confusion matrix for motorcycle make recognition.

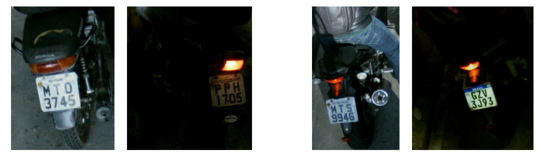
For motorcycle model recognition, a notable trend emerges: 14 of the 15 most frequent misclassifications occurred between models of the same make (see Fig. 6). Such errors often involve motorcycles designed for the same market segment or models that replaced earlier versions, reflecting the continuation of a brand's design language over time. The only exception among the top 15 errors involving models from different makes was motorcycles that competed in the same category.



Yamaha Factor Yamaha Fazer

Fig. 6. Examples of the two motorcycle models most frequently misclassified: Yamaha Factor (left) and Yamaha Fazer (right). Their overlapping visual features and shared market segment highlight the challenges of distinguishing between models from the same manufacturer.

A promising pattern involves low-light and nighttime images, which make up about 20% of both the *MotorcycleMake* and *MotorcycleModel* datasets. These images account for nearly one-third of all classification errors, indicating increased difficulty under such conditions. This is expected because the datasets primarily contain limited rear views of motorcycles, showing rear lamps (brake and turn signals) and part of the rear-lateral area. The classifier likely depends on these regions for recognition[2], but in low-light or nighttime images, visibility is mostly limited to the rear brake light, making identification more challenging (examples are shown in Fig. 7).



Daytime Nighttime Daytime Nighttime
Honda Pop-100 Kawasaki Ninja

Fig. 7. Comparison of motorcycle images captured in daytime and nighttime conditions. Each pair shows the same make and model, labeled below. Motorcycles are correctly recognized in daytime images but misclassified at nighttime. Daytime images provide clearer visual details, while nighttime views are mostly limited to rear brake lights, increasing classification difficulty under low-light conditions.

We also conducted a preliminary analysis to assess whether the presence of cargo boxes impacted classification performance but found no clear pattern. This may be due to the exclusion of heavily occluded motorcycles from the dataset. However, this issue warrants further attention, as cargo boxes are commonly used by riders in some countries (e.g., Brazil) and can partially obscure key motorcycle features.

V. Conclusions

This work-in-progress study investigates Fine-Grained Motorcycle Classification (FGMC) within a real-world Intelligent Transportation System (ITS) scenario. We evaluated seven deep learning architectures on motorcycle make and model recognition tasks under two training protocols, with an in-depth analysis of the best-performing model (EfficientNet-V2) to identify classification patterns and common errors. This study lays a foundation for FGMC solutions and calls for the inclusion of motorcycles in Fine-Grained Vehicle Classification (FGVC) research.

Our findings highlight three main points: (i) fine-grained classification of motorcycles at both make and model levels is achievable in ITS contexts and merits further focus within the FGVC community; (ii) severe class imbalance remains a critical challenge, emphasizing the need for improved data balancing or augmentation methods; and (iii) there is significant room for improvement, especially under challenging conditions like nighttime and low-light imagery.

Future work will expand the dataset to cover more diverse ITS environments beyond toll booths. Additionally, integrating FGMC with related ITS tasks such as license plate recognition

[2]We do not have direct evidence; however, given the limited rear view in the images, it is a plausible assumption. Further investigation would be valuable to confirm this and provide deeper insights.

offers a promising path to enhance overall vehicle identification accuracy. Finally, applying interpretability techniques to better understand the visual features driving classification decisions will further advance FGMC development.

ACKNOWLEDGMENTS

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)*, through the *Programa de Excelência Acadêmica (PROEX)* - Finance Code 001, and in part by the *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)* (# 315409/2023-1). We thank the support of NVIDIA Corporation with the donation of the Quadro RTX 8000 GPU used for this research.

REFERENCES

[1] S. H. Tan, J. H. Chuah, C.-O. Chow, and J. Kanesan, "Cross-granularity network for vehicle make and model recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 26, pp. 5782–5791, 2025.

[2] S. Wolf, D. Loran, and J. Beyerer, "Knowledge-distillation-based label smoothing for fine-grained open-set vehicle recognition," in *IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW)*, 2024, pp. 330–340.

[3] 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.

[4] V. Nascimento *et al.*, "Toward advancing license plate super-resolution in real-world scenarios: A dataset and benchmark," *Journal of the Brazilian Computer Society*, vol. 1, no. 31, pp. 435–449, 2025.

[5] L. Yang, P. Luo, C. C. Loy, and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 3973–3981.

[6] J. Krause, J. Deng, M. Stark, and L. Fei-Fei, "Collecting a large-scale dataset of fine-grained cars," https://ai.stanford.edu/ jkrause/papers/fgvc13.pdf, 2013.

[7] J. Masula, R. Tingga, J. A. Salido, J. C. Española, and G. K. Gillesania, "Motorcycle recognition system using convolutional neural network," *Journal of Innovative Technology Convergence*, vol. 6, no. 3, 2024.

[8] B. Hu, J.-H. Lai, and C.-C. Guo, "Location-aware fine-grained vehicle type recognition using multi-task deep networks," *Neurocomputing*, vol. 243, pp. 60–68, 2017.

[9] J. E. Espinosa, S. A. Velastin, and J. W. Branch, "Motorcycle classification in urban scenarios using convolutional neural networks for feature extraction," in *International Conference of Pattern Recognition Systems (ICPRS)*, 2017, pp. 1–6.

[10] P. K. Khoba, C. Parikh, C. Jawahar, R. K. Sarvadevabhatla, and R. Saluja, "A fine-grained vehicle detection (FGVD) dataset for unconstrained roads," in *Indian Conference on Computer Vision, Graphics and Image Processing*, 2023.

[11] S. M. Mansoor Roomi, V. Lokesh, S. Mahadevan G, and K. Priya, "A transfer learning approach for identification of two-wheeler brand and model," *Computación y Sistemas*, vol. 28, no. 2, pp. 543–551, 2024.

[12] Associação Brasileira dos Fabricantes de Motocicletas, Ciclomotores, Motonetas, Bicicletas e Similares, "Two wheel sector data 2024," https://abraciclo.com.br/wp-content/uploads/2025/04/Dados-do-Setor-de-Duas-Rodas-2025.pdf, 2025, accessed: 2025-08-10.

[13] Fantástico, "Fantástico apresenta raio-x inédito com o mapa dos roubos e furtos de moto em são paulo," https://g1.globo.com/fantastico/noticia/2025/04/20/fantastico-apresenta-raio-x-inedito-com-o-mapa-dos-roubos-e-furtos-de-moto-em-sao-paulo.ghtml, 2025, accessed: 2025-08-10.

[14] L. Wojcik, G. E. Lima, V. Nascimento, E. Nascimento Jr., R. Laroca, and D. Menotti, "LPLC: A dataset for license plate legibility classification," *Conference on Graphics, Patterns and Images*, pp. 1–6, 2025.

[15] X. Clady, P. Negri, M. Milgram, and R. Poulenard, "Multi-class vehicle type recognition system," in *Artificial Neural Networks in Pattern Recognition*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 228–239.

[16] V. Petrovic and T. Cootes, "Analysis of features for rigid structure vehicle type recognition," in *British Machine Vision Conference (BMVC)*, vol. 2, 2004.

[17] R. Laroca, E. V. Cardoso, D. R. Lucio, V. Estevam, and D. Menotti, "On the cross-dataset generalization in license plate recognition," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, Feb 2022, pp. 166–178.

[18] Q. Liu, Y. Liu, S.-L. Chen, T.-H. Zhang, F. Chen, and X.-C. Yin, "Improving multi-type license plate recognition via learning globally and contrastively," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 9, pp. 11 092–11 102, 2024.

[19] G. E. Lima, R. Laroca, E. Santos, E. Nascimento Jr., and D. Menotti, "Toward enhancing vehicle color recognition in adverse conditions: A dataset and benchmark," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Sept 2024, pp. 1–6.

[20] V. Nascimento *et al.*, "Enhancing license plate super-resolution: A layout-aware and character-driven approach," *Conference on Graphics, Patterns and Images (SIBGRAPI)*, pp. 1–6, 2024.

[21] R. Laroca, L. A. Zanlorensi, V. Estevam, R. Minetto, and D. Menotti, "Leveraging model fusion for improved license plate recognition," in *Iberoamerican Congress on Pattern Recognition*, Nov 2023, pp. 60–75.

[22] Ultralytics, "Ultralytics yolo11," https://docs.ultralytics.com/pt/models/yolo11/, 2025, accessed: 2025-08-10.

[23] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243–9275, 2023.

[24] J. Terven, D.-M. Córdova-Esparza, and J.-A. Romero-González, "A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS," *Machine Learning and Knowledge Extraction*, vol. 5, no. 4, pp. 1680–1716, 2023.

[25] R. Laroca, M. dos Santos, and D. Menotti, "Improving small drone detection through multi-scale processing and data augmentation," in *International Joint Conference on Neural Networks (IJCNN)*, 2025.

[26] V. Montoia, "Lista: veja quais foram as motos mais vendidas do brasil até julho 2025," https://g1.globo.com/carros/motos/noticia/2025/08/06/lista-veja-quais-foram-as-motos-mais-vendidas-do-brasil-ate-julho.ghtml, 2025, accessed: 2025-08-10.

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

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

[29] 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.

[30] Z. Liu, H. Hu, Y. Lin, Z. Yao, Z. Xie, Y. Wei, J. Ning, Y. Cao, Z. Zhang, L. Dong, F. Wei, and B. Guo, "Swin transformer v2: Scaling up capacity and resolution," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 11 999–12 009.

[31] 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. 1–22.

[32] Ultralytics, "YOLOv11 Image Classification," https://docs.ultralytics.com/tasks/classify/, 2025, accessed: 2025-08-10.

[33] ——, "Data augmentation using Ultralytics YOLO," https://docs.ultralytics.com/pt/guides/yolo-data-augmentation/, 2025, accessed: 2025-08-10.

[34] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "Randaugment: Practical automated data augmentation with a reduced search space," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 3008–3017.

[35] R. Laroca, V. Estevam, A. S. Britto Jr., R. Minetto, and D. Menotti, "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.

[36] F. Wilcoxon, *Individual Comparisons by Ranking Methods*. New York, NY: Springer New York, 1992, pp. 196–202. [Online]. Available: https://doi.org/10.1007/978-1-4612-4380-9_16

[37] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, p. 1–30, Dec. 2006.