Improving Vehicle Identification Through Advanced Fine-Grained Vehicle Classification

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 identification plays a crucial role in Intelligent Transportation Systems, impacting areas such as toll collection, vehicle access control, and criminal forensics. Despite recent strides in Automatic License Plate Recognition (ALPR) research, real-world scenarios still pose significant challenges. This work explores potential enhancements in vehicle identification systems by integrating modules such as ALPR with Fine-Grained Vehicle Classification (FGVC), which categorizes vehicles based on attributes such as type, make, model, and year. Our study focuses on advancing FGVC, particularly vehicle type classification. We investigate selective prediction, a technique that allows models to discard uncertain predictions, and examine superclass methods, including a novel online superclass approach that operates solely during the test phase. We trained and evaluated four deep learning models using a dataset adapted from a widely adopted ALPR dataset. The results demonstrate that both superclass methods and selective prediction improve classification accuracy, with the combination of online superclass and selective prediction delivering the best performance. Future research will focus on integrating these enhancements into ALPR systems to determine how FGVC can further enhance their capabilities.**

I. Introduction

Vehicle identification is an important component of Intelligent Transportation Systems, with various applications ranging from toll collection and vehicle access control to criminal forensics [1]–[4]. Despite recent advancements that have led to high recognition rates in Automatic License Plate Recognition (ALPR), significant challenges remain in real-world scenarios [5]–[7]. For instance, the Military Police of Paraná in Brazil has reported frequent inaccuracies in their ALPR system during actual deployments. These issues underscore the need for more robust vehicle identification methods, as inaccuracies can severely undermine the reliability and effectiveness of these systems in critical situations.

To improve vehicle identification methods, we propose integrating ALPR techniques, which have become increasingly common in recent years, with Fine-Grained Vehicle Classification (FGVC). FGVC categorizes vehicle images based on specific attributes such as type, make, model, and year [8]–[10]. This integration offers two key advantages. First, by cross-referencing vehicle attributes with documentation, it can effectively detect ALPR errors, reducing false positives. Second, FGVC can help narrow the search space for license plates by utilizing vehicle attributes, which is particularly useful in cases where license plates are low-quality or obscured.

The FGVC field is inherently challenging due to the need for a precise understanding of vehicle appearance, as many vehicles can look remarkably similar. However, research exploring deep learning models, such as Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), has achieved classification accuracies exceeding 90% [11]–[13]. Despite these advancements, certain aspects that could benefit the proposed integration remain underexplored. Research on FGVC often assumes well-controlled conditions [12], [14], whereas ALPR systems typically operate in more challenging environments. This highlights the need to improve the robustness of FGVC systems to ensure reliable performance in real-world, surveillance scenarios.

One way to enhance classification performance is through selective prediction, where machine learning models reject uncertain results, reducing error rates and flagging data for manual verification [15], [16]. Another strategy involves using a superclass approach, which reduces the number of classes by clustering them based on specific criteria [17]. While these methods have not yet been explored for FGVC, they have shown effectiveness in improving classification tasks in other contexts [17]–[20].

In light of this, this work-in-progress outlines the initial steps in applying selective prediction and superclass approaches to vehicle type classification. Vehicle type was chosen because it offers a fundamental level of classification with fewer categories, which streamlines the preliminary analysis. Despite its apparent simplicity, vehicle type information is valuable for identifying errors in license plate recognition. For example, distinguishing between cars, trucks, and motorcycles can aid in manual verification to validate recognition results.

Four deep learning models were evaluated for classifying vehicle types using an adapted version of the RodoSol-ALPR dataset [21]. The best-performing model was further analyzed under various scenarios: (i) employing a Softmax Response Rejection (SRR) scheme; (ii) training and evaluating with superclasses; (iii) applying a naive superclass scheme solely during evaluation, referred to as online superclass; and (iv) combining online superclass with selective prediction.

The remainder of this work is organized as follows. Section II briefly reviews related works. Section III describes the data preparation process. Section IV details the methodology used in the experiments. Section V presents the results obtained. Finally, Section VI summarizes our findings and suggests directions for future research.

II. Related Work

Extensive research on FGVC has yielded many promising results [8], [11], [12]. In this work-in-progress, we specifically focus on studies related to vehicle type classification.

Ferryman et al. [22] developed a 3D model to recover vehicle pose and structure information, enabling differentiation between vehicle types. Jolly et al. [23] proposed a segmentation algorithm integrating motion and edge data with deformable templates, which was then used for classification through template matching. Lai et al. [24] conducted classification using a visual-based approach that estimates the width, length, and height of vehicles. Wu et al. [25] introduced a parameterized edge model combined with a multilayer perceptron to capture vehicle structure and perform classification. Ma et al. [26] enhanced feature discriminability by associating edge points with SIFT-based descriptors and applied clustering techniques for classification.

Dong et al. [27] were pioneers in applying CNNs for vehicle type classification, using a sparse Laplacian filter learning method with large amounts of unlabeled data. Following this, Hu et al. [28] developed a multi-task CNN that localizes vehicles and performs fine-grained classification, leveraging collaborative feature learning to handle subtle inter-class variations. Wang et al. [29] explored Faster R-CNN for both vehicle detection and classification. Shvai et al. [30] employed a Gradient Boosting classifier to integrate the outputs from a CNN classifier and an optical sensor-based classifier. Kim [14] applied YOLOv4 for vehicle detection and used a pre-trained ResNet-50 for vehicle type classification.

These studies demonstrate the shift from traditional handcrafted methods to advanced deep learning techniques in vehicle type classification, marking significant progress in the field. However, there is a noticeable gap in the literature regarding the integration of vehicle type classification with ALPR systems. While FGVC research has shown promising results, it is often conducted under well-controlled conditions, such as single viewpoints and stable lighting. These controlled environments do not fully represent the diverse and challenging conditions encountered in real-world scenarios where ALPR systems are commonly applied, potentially complicating the integration of these two fields.

III. Data Preparation

This section outlines the data preparation process for the experimental research. The RodoSol-ALPR dataset [21] was selected for model training and evaluation. This dataset comprises 20,000 images captured by static cameras at toll booths on a Brazilian highway (see Fig. 1). We chose this dataset due to its: (i) widespread use in ALPR research [6], [31], [32]; (ii) representation of real-world intelligent transportation systems scenarios; and (iii) accurate license plate annotations, which enable the derivation of vehicle attributes for FGVC and support research that integrates both domains.

Fig. 1. Samples from the RodoSol-ALPR dataset [21]. These images illustrate the diversity of vehicle types and varying lighting conditions captured by the dataset. The original images have been slightly resized for better viewing.

Similar to [33], we standardized the images by cropping the vehicles to remove any background, preparing the dataset for the proposed classification task. As the RodoSol-ALPR dataset does not provide vehicle bounding box annotations, we applied YOLOv8 [34] for vehicle detection. To preserve data integrity and accurately reflect real-world conditions, we refrained from applying noise reduction or image enhancement techniques.

The next step involved a manual review and filtering process to select images suitable for vehicle type classification. We first grouped images of the same vehicle based on the license plate annotations, recognizing that a vehicle might appear multiple times on different days or at different times. We retained samples that exhibited significant variations in lighting, pose, or other distinguishing attributes. Additionally, we excluded images where the vehicle was heavily occluded; for example, images showing only the vehicle bumper were discarded, as they are inadequate for accurate vehicle type classification.

Finally, we carried out the annotation process. Each image was initially linked to its corresponding vehicle license plate, enabling us to automatically retrieve vehicle information from Brazil's National Traffic Secretariat (SENATRAN) database. We identified 12,785 unique license plates, but 424 of these lacked complete data. Consequently, we manually labeled the vehicle types for the remaining entries.

After completing the preparation process, the dataset, hereinafter referred to as *Vehicle-Type*[1], comprised 17,393 images categorized into eleven classes: bus, car, minibus, motorcycle, pickup, scooter, Sports Utility Vehicle (SUV), subcompact SUV, tractor-trailer, tricycle, and truck. Fig. 2 shows the final class distribution, which exhibits a long-tailed pattern typical of Brazilian traffic scenarios, where motorcycles and cars are more prevalent than other vehicle types [35].

Fig. 3 showcases examples of images from each class within the dataset, highlighting the close relationships between certain classes. For example, the scooter and motorcycle classes feature similar vehicles with subtle differences. Despite these similarities, all classes were maintained separately to ensure the data's fidelity, as recorded in the SENATRAN database.

[1] The list of selected images from the RodoSol-ALPR dataset, along with their corresponding vehicle type annotations and our division into training, test and validation subsets, will be provided upon request.

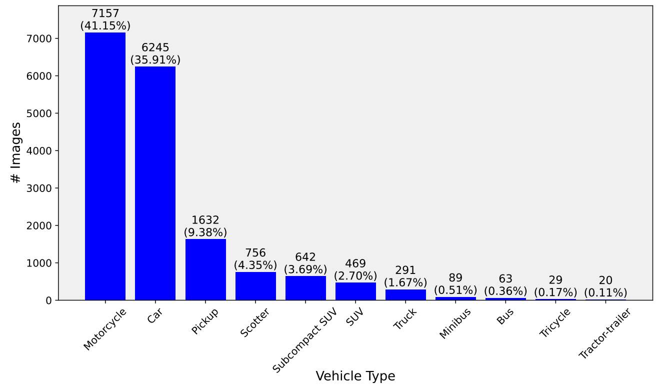

Fig. 2. Distribution of vehicle types in the Vehicle-Type dataset.

Fig. 3. Examples from the Vehicle-Type dataset, obtained after applying the data preparation process on the RodoSol-ALPR dataset [21]. Each image is annotated with its corresponding vehicle type, displayed above the image. Here, all images have been resized to a uniform size for better viewing.

IV. Methodology

This section describes the experimental research conducted to investigate the impact of selective prediction and superclass approaches on the FGVC problem. The initial experiment, referred to as the baseline experiment, involved training four models for vehicle type classification (Section IV-A). The best-performing model was then subjected to further analysis under various scenarios: (i) retraining and evaluating with the superclass approach and the proposed online superclass scheme (Section IV-B); (ii) assessing with the SRR selective prediction method (Section IV-C); and (iii) evaluating by combining online superclass with selective prediction (Section IV-D).

Each experiment was conducted five times with different dataset splits, and the results are reported as the average outcomes. The metrics include Top-1 accuracy, Top-2 accuracy, precision, recall, and F1-score. Subsequent experiments focus on analyzing the correct predictions of the best-performing model, with results reported in terms of Top-1 and Top-2 accuracy. All metrics are presented using macro averaging.

The dataset was split into training, validation, and test subsets in an $8{:}1{:}1$ ratio, maintaining the original class distribution percentages in each subset. When exact class percentages did not allow for a precise division, any surplus images were randomly assigned to either the validation or test subsets.

A. Baseline Experiment

The models explored in this study are ResNet-34 [36], MobileNetV3 [37], EfficientNetV2 [38], and ViT b16 [39]. They were selected for their widespread application in image classification tasks, including fine-grained classification research [40]–[42], as well as their availability of pre-trained implementations across various frameworks, which enhances reproducibility. We applied a transfer learning approach for training: each model was initialized with pre-trained weights from ImageNet, and only the final fully connected layer was modified to match the prediction classes. During training, adjustments were made exclusively to these final layers.

The Adam optimizer was used with $\beta_1 = 0.9$, $\beta_2 = 0.999$, a batch size of 128, weight decay of 10^{-5}, and an initial learning rate of 10^{-4}. A learning rate reduction scheme was configured with a patience of 10 epochs and a decrease factor of 0.1. The training was capped at 400 epochs, with early stopping applied if no improvement was observed over 15 epochs. Cross-entropy loss was employed as the loss function.

During training and evaluation, all images were resized to 224×224 pixels to meet the models' input requirements. To increase data variability, the following transformations were applied to each training batch at every epoch: (i) rotations up to $180°$, scaling between 0.9 and 1.3, and shearing up to $30°$, each with a probability of 50%; (ii) random changes to image brightness and contrast within a range of 0.2, with a probability of 30%; (iii) blurring using a generalized normal filter with randomly selected parameters, with a probability of 40%; (iv) A random 72×72 pixel section is replaced with random noise, with a probability of 25%; (v) All images were normalized using the mean and standard deviation from ImageNet.

Two training protocols were adopted: *(p1)* training with data augmentation alone; and *(p2)* training with the oversampling of minority classes to balance the dataset distribution, which enhances performance on imbalanced datasets [43]. The second protocol increases the representation of minority classes across batches, generating synthetic data through the data augmentation techniques mentioned above.

B. Superclass and Online Superclass

The superclass experiment aims to establish a higher-level class distribution for classification tasks [17], [19]. This is achieved through semantic clustering, which groups images of vehicles with similar characteristics into a common superclass. Table I presents the mapping of individual classes to their respective superclasses and the resulting class distribution. Thus, the superclass experiment involves replicating the baseline methodology, but using the classes as defined by this mapping.

The online superclass experiment assesses whether retraining the best model from the baseline experiment is necessary

TABLE I
MAPPING OF ORIGINAL CLASSES TO SUPERCLASSES BASED ON SIMILAR VEHICLE CHARACTERISTICS, ALONG WITH THE RESULTING DISTRIBUTION OF CLASSES IN THE NEW DATASET.

Original Class	Superclass	Images
Motorcycle Scooter Tricyle	Motorcycle	7,942
Car	Car	6,245
Pick-up SUV Subcompact SUV	SUV	2,743
Tractor-trailer Truck	Truck	311
Bus Minibus	Bus	152

to achieve the results of the superclass experiment. Instead of retraining, we incorporate superclass mapping only during the evaluation phase. For instance, if a data sample is labeled as "motorcycle" and the model predicts "scooter," the prediction is deemed correct because "scooter" belongs to the same superclass as "motorcycle" (see Table I).

C. Selective Prediction

The naive Softmax Response Rejection (SRR) method [15], [20] is applied using the best-performing model from the baseline experiment. This method is chosen for its simplicity in determining whether to accept or reject predictions, as it requires no modifications to the model architecture or training process. It relies on the probability distribution from the softmax layer of the classification network. A threshold is set, and predictions are accepted if the highest softmax probability surpasses this threshold; otherwise, they are rejected and excluded from the evaluation metrics.

This experiment thoroughly examined selective prediction by testing nine threshold values, ranging from 0.1 to 0.9 in increments of 0.1. Additionally, it tracked extra metrics, including the number of rejected images and the originally correct predictions that were incorrectly rejected, both as absolute numbers and percentages from the test subset. These metrics provide a deeper understanding of the impact of the SRR approach on the FGVC task.

D. Online Superclass and Selective Prediction

To further investigate the effects of both the superclass and selective prediction approaches, a final experiment was conducted that combines the methodologies from Sections IV-B and IV-C. The online superclass method was selected for this experiment because it operates exclusively during the evaluation phase, similar to the selective prediction method.

V. PARTIAL RESULTS

Table II presents the results of the baseline experiment for each model on the evaluated dataset, following the two training protocols described in Section IV-A. Notably, models trained using protocol (p2) achieved higher accuracy compared to those trained with protocol (p1). This outcome was expected, as the second protocol enhanced accuracy for minority classes, thereby increasing the overall macro average. However, this approach also led to a decrease in the models' precision.

TABLE II
GLOBAL METRICS REACHED BY ALL MODELS ON THE VEHICLE-TYPE DATASET, AVERAGED OVER FIVE RUNS. PROTOCOL *(p2)* INCORPORATES OVERSAMPLING OF MINORITY CLASSES, WHEREAS *(p1)* DOES NOT.

Protocol	Model	Top-1	Top-2	Precision	Recall	F1
(p1)	ViT b16	**65.9%**	**88.1%**	**75.3%**	**65.9%**	**69.1%**
	ResNet 34	58.2%	80.4%	76.6%	58.2%	64.0%
	EfficientNetV2	50.6%	77.2%	69.2%	50.6%	56.1%
	MobileNetV3	61.7%	78.8%	73.5%	61.7%	65.5%
(p2)	ViT b16	**78.2%**	**92.0%**	**65.9%**	**78.2%**	**70.2%**
	ResNet 34	74.7%	89.4%	53.1%	74.7%	58.0%
	EfficientNetV2	73.7%	87.7%	50.1%	73.7%	55.8%
	MobileNetV3	70.3%	86.6%	51.5%	70.3%	57.1%

Focusing on correct predictions, the ViT b16 model emerged as the top performer, achieving 78% Top-1 and 92% Top-2 accuracy when trained under protocol *(p2)*. These results highlight the model's strong generalization capabilities, even in the challenging FGVC scenario. Consequently, the ViT b16 trained with protocol *(p2)* was selected for further analysis to explore potential improvements in classification accuracy through the superclass and selective prediction approaches.

Table III compares the baseline results with those obtained using the ViT b16 model across the two superclass approaches. Both superclass methods showed higher accuracy than the baseline model. This improvement is attributed to the reduced task complexity, achieved by clustering similar classes together based on their visual characteristics. The normalized average confusion matrix from the best-performing model in the baseline experiment (Fig. 4) shows that similar vehicle types exhibited noticeable misclassifications. Thus, grouping visually similar classes can minimize incorrect predictions.

TABLE III
GLOBAL TOP-1 AND TOP-2 ACCURACY VALUES REACHED BY VIT B16 ON THE VEHICLE-TYPE DATASET USING THE SUPERCLASS AND ONLINE SUPERCLASS APPROACHES, AVERAGED OVER FIVE RUNS.

Method	Top-1	Top-2
Baseline	78.2%	92.0%
Superclass	87.8%	**98.1%**
Online Superclass	**88.0%**	96.7%

Additionally, it is worth noting that the Top-1 accuracy difference between the two superclass methods is minimal. Further analysis using a paired t-test at a 5% significance level shows no statistically significant difference between the methods. Nevertheless, the online superclass method is preferred because it enables the model to retain the ability to predict the original classes.

Table IV presents the results reached by the ViT b16 model (from the baseline experiment) when applying selective prediction with SRR across different minimum confidence thresholds. Naturally, higher thresholds lead to the rejection

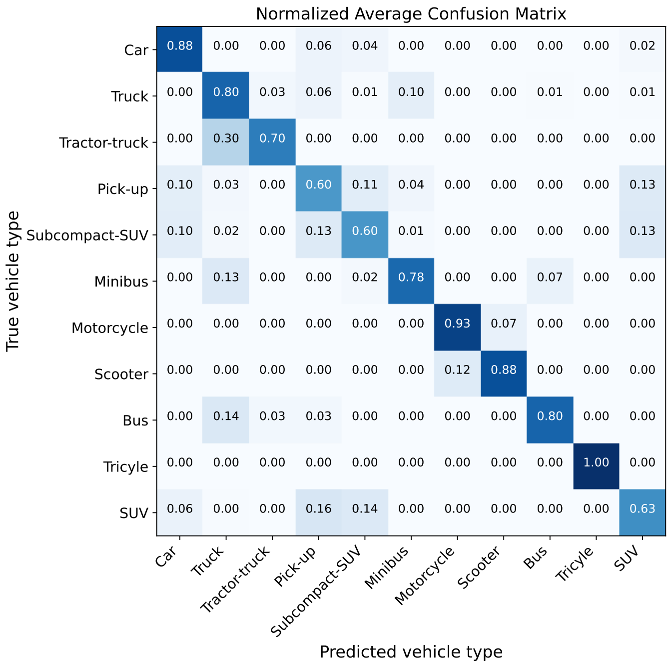

Fig. 4. Normalized confusion matrix illustrating the performance of the ViT b16 model trained with data augmentation and oversampling techniques, which achieved the best results on the baseline experiment.

of more images. The range of minimum confidence thresholds produced rejection rates from 0.4% (with a threshold of 0.3) to 55% (with a threshold of 0.9) of the total test subset.

TABLE IV
GLOBAL TOP-1 AND TOP-2 ACCURACY VALUES ACHIEVED BY VIT B16 ON THE VEHICLE-TYPE DATASET USING SOFTMAX RESPONSE REJECTION, AVERAGED OVER FIVE RUNS. REJECTION RATES ARE SHOWN AS BOTH ABSOLUTE NUMBERS AND PERCENTAGES.

Minimum Confidence	Rejected Images	Correct Predictions Incorrectly Rejected	Top-1	Top-2
0.1	0 / 0.0%	0 / 0.0%	78.2%	92.0%
0.2	0 / 0.0%	0 / 0.0%	78.2%	92.0%
0.3	7 / 0.4%	2 / 0.1%	78.5%	92.4%
0.4	71 / 4.1%	28 / 1.6%	80.5%	94.0%
0.5	192 / 11.0%	92 / 5.3%	83.4%	95.0%
0.6	348 / 20.0%	192 / 11.0%	86.9%	95.6%
0.7	503 / 28.9%	308 / 17.7%	86.1%	92.3%
0.8	692 / 39.8%	472 / 27.4%	86.9%	92.9%
0.9	968 / 55.7%	730 / 42.0%	85.5%	90.3%

Although raising the threshold improves accuracy compared to the baseline, a significant issue arises: as the rejection rate increases, so does the number of correct predictions that are incorrectly rejected. In the worst-case scenario (threshold of 0.9), 42% of initially correct classifications were rejected. This results in a plateau in accuracy improvement as the threshold increases. For confidence thresholds above 0.6, accuracy gains level off and may even decline compared to lower thresholds Therefore, balancing accuracy and rejection rate is crucial for optimizing results. At a threshold of 0.5, baseline accuracy is improved with an 11.0% rejection rate, while minimizing the loss of correct predictions to 5.3% of them.

Table V shows the results of combining selective prediction with the online superclass method. This improved the baseline results without requiring retraining or changes to the best-performing model architecture. Remarkably, with a confidence threshold of 0.4, we achieved a rejection rate of 4.1% and Top-1/Top-2 accuracy rates of 90.1% and 98.3%, respectively.

TABLE V
GLOBAL TOP-1 AND TOP-2 ACCURACY VALUES REACHED BY VIT B16 ON THE VEHICLE-TYPE DATASET USING SRR AND ONLINE SUPERCLASS, AVERAGED OVER FIVE RUNS. REJECTION RATES ARE SHOWN AS BOTH ABSOLUTE NUMBERS AND PERCENTAGES.

Minimum confidence	Rejected Images	Correct Predictions Incorrectly Rejected	Top-1	Top-2
0.1	0 / 0.0%	0 / 0.0%	88.0%	96.7%
0.2	0 / 0.0%	0 / 0.0%	88.0%	96.7%
0.3	7 / 0.4%	4 / 0.2%	88.4%	97.1%
0.4	71 / 4.1%	45 / 2.6%	90.1%	98.3%
0.5	192 / 11.0%	129 / 7.4%	91.3%	98.7%
0.6	348 / 20.0%	260 / 15.0%	92.9%	98.9%
0.7	503 / 28.9%	399 / 22.9%	93.4%	99.1%
0.8	692 / 39.8%	580 / 33.4%	93.6%	99.2%
0.9	968 / 55.7%	850 / 48.9%	94.3%	99.3%

In conclusion, online superclass and selective prediction approaches led to notable improvements over the baseline. Top-1 accuracy increased from 78.2% to 87.8% with online superclass and to 88.9% with selective prediction. Similarly, Top-2 accuracy rose from 92.0% to 98.1% and 96.7%, respectively. By combining these techniques, final accuracy rates reached 90.1% for Top-1 and 98.3% for Top-2, with a rejection rate of $\approx 4.1\%$ of total predictions. These results demonstrate the effectiveness of the studied methods.

VI. CONCLUSIONS

This work represents an initial exploration of enhancing vehicle identification systems through the integration of modules such as Automatic License Plate Recognition (ALPR) and Fine-Grained Vehicle Classification (FGVC). Our work-in-progress study identified areas within existing FGVC literature that could benefit from this integration and introduced preliminary methods that incorporate both superclass and selective prediction approaches into vehicle type classification.

The experiments showed that both selective prediction and superclass methods can improve overall classification accuracy. It is noteworthy that adapting a model to evaluate predictions based on a superset of classes can yield results similar to training a model from scratch with these superclasses. Additionally, balancing rejection rates with accuracy improvements prediction remains an area for future research.

In conclusion, the combination of the studied methods shows promise for advancing FGVC and integrating it with ALPR systems. Future research should focus on several key areas: (i) improving the superclass method by employing automatic clustering techniques to map original classes to superclasses; (ii) refining the selective prediction scheme using more robust techniques, such as confidence calibration and learning with rejection; (iii) developing a combined ALPR and FGVC system to assess how vehicle attribute classification can further enhance license plate recognition accuracy.

ACKNOWLEDGMENTS

This study was financed in part by the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)* - 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] R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, and D. Menotti, "An efficient and layout-independent automatic license plate recognition system based on the YOLO detector," *IET Intelligent Transport Systems*, vol. 15, no. 4, pp. 483–503, 2021.

[2] 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.

[3] X. Ke, G. Zeng, and W. Guo, "An ultra-fast automatic license plate recognition approach for unconstrained scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, pp. 5172–5185, 2023.

[4] 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.

[5] V. Nascimento *et al.*, "Combining attention module and pixel shuffle for license plate super-resolution," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2022, pp. 228–233.

[6] V. Nascimento *et al.*, "Super-resolution of license plate images using attention modules and sub-pixel convolution layers," *Computers & Graphics*, vol. 113, pp. 69–76, 2023.

[7] C. He, D. Wang, Z. Cai, J. Zeng, and F. Fu, "A vehicle matching algorithm by maximizing travel time probability based on automatic license plate recognition data," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 8, pp. 9103–9114, 2024.

[8] 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.

[9] F. Oliveira, A. Macena, O. Kamel, W. Souza, N. Freitas, and T. Vinuto, "Fine-grained cars recognition using deep convolutional neural networks," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2022, pp. 240–245.

[10] D. M. Kuhn and V. P. Moreira, "BRCars: A dataset for fine-grained classification of car images," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, 2021, pp. 231–238.

[11] H. Lu, M. Han, C. Wang, and J. Cheng, "AMLNet: Attention multi-branch loss CNN models for fine-grained vehicle recognition," *IEEE Transactions on Vehicular Technology*, vol. 73, no. 1, pp. 375–384, 2024.

[12] 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.

[13] K. Valev *et al.*, "A systematic evaluation of recent deep learning architectures for fine-grained vehicle classification," in *Pattern Recognition and Tracking*, vol. 10649, 2018, p. 1064902.

[14] J. Kim, "Deep learning-based vehicle type and color classification to support safe autonomous driving," *Applied Sciences*, vol. 14, no. 4, 2024.

[15] Y. Geifman and R. El-Yaniv, "Selective classification for deep neural networks," in *International Conference on Neural Information Processing Systems (NeurIPS)*, 2017, p. 4885–4894.

[16] C. Cortes *et al.*, "Learning with rejection," in *Algorithmic Learning Theory*. Springer International Publishing, 2016, pp. 67–82.

[17] Y. Zhou, Q. Hu, and Y. Wang, "Deep super-class learning for long-tail distributed image classification," *Pattern Recognition*, vol. 80, pp. 118–128, 2018.

[18] K. Hendrickx, L. Perini, D. Van der Plas, W. Meert, and J. Davis, "Machine learning with a reject option: A survey," *Machine Learning*, vol. 113, no. 5, pp. 3073–3110, 2024.

[19] Z. Gan *et al.*, "Superclass learning with representation enhancement," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2023, pp. 24 060–24 069.

[20] X.-Y. Zhang *et al.*, "A survey on learning to reject," *Proceedings of the IEEE*, vol. 111, no. 2, pp. 185–215, 2023.

[21] 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.

[22] J. M. Ferryman, A. D. Worrall, G. D. Sullivan, and K. D. Baker, "A generic deformable model for vehicle recognition," in *British Machine Vision Conference (BMVC)*, 1995, p. 127–136.

[23] M.-P. Dubuisson Jolly, S. Lakshmanan, and A. Jain, "Vehicle segmentation and classification using deformable templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 293–308, 1996.

[24] A. Lai, G. Fung, and N. Yung, "Vehicle type classification from visual-based dimension estimation," in *IEEE Intelligent Transportation Systems Conference (ITSC)*, 2001, pp. 201–206.

[25] W. Wu, Z. QiSen, and W. Mingjun, "A method of vehicle classification using models and neural networks," in *IEEE Vehicular Technology Conference*, 2001, pp. 3022–3026.

[26] X. Ma and W. Grimson, "Edge-based rich representation for vehicle classification," in *IEEE International Conference on Computer Vision (ICCV)*, 2005, pp. 1185–1192.

[27] Z. Dong, Y. Wu, M. Pei, and Y. Jia, "Vehicle type classification using a semisupervised convolutional neural network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247–2256, 2015.

[28] 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.

[29] X. Wang *et al.*, "Real-time vehicle type classification with deep convolutional neural networks," *Journal of Real-Time Image Processing*, vol. 16, pp. 5–14, 2019.

[30] N. Shvai, A. Hasnat, A. Meicler, and A. Nakib, "Accurate classification for automatic vehicle-type recognition based on ensemble classifiers," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 3, pp. 1288–1297, 2020.

[31] R. Laroca, M. Santos, V. Estevam, E. Luz, and D. Menotti, "A first look at dataset bias in license plate recognition," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2022, pp. 234–239.

[32] 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*, pp. 1–11, 2024, early Access.

[33] 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, in Press.

[34] Ultralytics, "YOLOv8," 2023, accessed: 2024-07-28. [Online]. Available: https://github.com/ultralytics/ultralytics

[35] Ministério dos Transportes, "Frota nacional (junho de 2024)," https://www.gov.br/transportes/pt-br/assuntos/transito/conteudo-Senatran/frota-de-veiculos-2024, 2024, accessed: 2024-08-09.

[36] 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.

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

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

[39] 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.

[40] X.-S. Wei, Y.-Z. Song, O. Mac Aodha, J. Wu, Y. Peng, J. Tang, J. Yang, and S. Belongie, "Fine-grained image analysis with deep learning: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 12, pp. 8927–8948, 2022.

[41] L. Lu, P. Wang, and Y. Cao, "A novel part-level feature extraction method for fine-grained vehicle recognition," *Pattern Recognition*, vol. 131, no. 108869, p. 108869, 2022.

[42] A. Hassan, M. Ali, N. M. Durrani, and M. A. Tahir, "An empirical analysis of deep learning architectures for vehicle make and model recognition," *IEEE Access*, vol. 9, pp. 91 487–91 499, 2021.

[43] 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.