Car Classification using ResNet-18 on the CompCars Dataset

Daria Nikolaeva[†], Andrey Malikov[†], Nichita Gramatchi[†]

***Abstract*—Image classification is a crucial task in computer vision with applications in autonomous driving, traffic monitoring, and vehicle identification. In this work, we explore the use of convolutional neural networks (CNNs) to classify car manufacturers using the CompCars dataset, which consists of images captured from surveillance cameras.**

We implement different preprocessing techniques, including data augmentation, RGB input, and class imbalance handling. Our final model is based on ResNet-18, trained on a well-balanced dataset using early stopping.

Our results show that the model achieves 89.66% accuracy on the test set, demonstrating strong generalization to unseen images. We also analyze misclassifications using a confusion matrix to identify challenging cases. This study provides insights into CNN-based vehicle classification and lays the foundation for further improvements through fine-tuning and larger datasets.

***Index Terms*—Supervised Learning, Data augmentation, Classification, Convolutional Neural Networks, Residual Networks.**

I. Introduction

A. Background and Motivation

Deep learning has revolutionized computer vision and has become a fundamental tool in object classification, detection, and recognition. In particular, vehicle classification plays a crucial role in traffic monitoring, law enforcement, and intelligent transportation systems. Traditional methods based on handcrafted features and classical machine learning models struggle with large-scale real-world data due to variations in lighting conditions, viewing angles, and occlusions.

B. Problem Statement and Challenges

Unlike standard object recognition tasks, fine-grained car classification must distinguish visually similar manufacturers despite challenging real-world conditions. The CompCars dataset [1], one of the largest publicly available car datasets, provides both web-based and surveillance-based vehicle images. We focus exclusively on the surveillance subset, which consists of low-angle, real-world images captured by traffic cameras. This dataset poses several key challenges:

- **Class imbalance:** Some manufacturers (e.g., Volkswagen) are heavily overrepresented, while others (e.g., Infiniti) have very few samples.
- **Difficult viewing conditions:** Surveillance images often suffer from blur, extreme lighting conditions, occlusions, and limited feature visibility.
- **Fine-grained recognition complexity:** Different brands often share similar designs (e.g., BMW vs. Mercedes-Benz), making classification challenging.

C. Proposed Approach

In this work, we propose a deep learning-based solution using Convolutional Neural Networks (CNNs) to classify car manufacturers from surveillance images in the CompCars dataset. Specifically, we implement **ResNet-18**, a widely used deep learning architecture [2], and introduce the following enhancements:

- **Class imbalance handling:** We apply *Focal Loss* instead of traditional *Cross-Entropy Loss* to improve classification on underrepresented brands.
- **Data augmentation:** Targeted transformations (random flips, rotation, color jitter) are applied *only to rare classes* to prevent overfitting.
- **Fine-tuning ResNet-18**: We modify the fully connected layer to adapt to 67 car brands.
- **Early stopping:** This technique improves model generalization and prevent excessive training.
- **Confusion matrix and error analysis:** We analyze misclassified samples and identify patterns in errors.

D. Summary of Contributions

This work makes the following contributions:

- We develop a ResNet-18-based classification model for identifying car manufacturers from surveillance images.
- We handle class imbalance using Focal Loss and targeted data augmentation.
- We implement early stopping to optimize training.
- We analyze model misclassifications and provide insights for further improvements.

E. Paper Organization

The rest of the report is structured as follows: in Section II we discuss related work. Section III describes the dataset and preprocessing techniques. Section IV presents our CNN model and training strategy. Section V reports experimental results and evaluation metrics. Finally, Section VI concludes the paper and suggests future improvements.

II. Related Work

Car classification is a well-established problem in computer vision, gaining increasing relevance due to applications in autonomous driving, traffic monitoring, and surveillance. The introduction of large-scale datasets has enabled significant

[†] Department of Information Engineering, University of Padova, email: {name.surname}@studenti.unipd.it

progress in this field. Among them, the CompCars dataset [1] is one of the most comprehensive, containing both web and surveillance images. While web images provide high-quality, well-lit samples, surveillance images introduce real-world challenges such as low resolution, occlusions, and varying lighting conditions. Although CNN-based methods have demonstrated high accuracy on web images, the surveillance subset remains underexplored, motivating our focus on this subset.

Deep learning, particularly Convolutional Neural Networks (CNNs), has become the dominant approach for fine-grained vehicle classification. Studies such as [3] highlight the importance of deep architectures like ResNet, VGG, and Inception, which are capable of extracting hierarchical features essential for distinguishing visually similar car models. Given the trade-off between accuracy and computational efficiency, we adopt ResNet-18, a well-balanced architecture offering strong feature extraction while maintaining reasonable computational costs.

One of the key challenges in real-world vehicle classification is class imbalance, where frequent car brands (e.g., Volkswagen, Toyota) dominate datasets, leading to biased models. Previous research [4] has tackled this issue using hierarchical classification, which first predicts the vehicle type before classifying the manufacturer and model. However, such multi-stage approaches increase computational complexity. Instead, we address class imbalance using Focal Loss, which prioritizes hard-to-classify samples, and targeted data augmentation, which selectively enhances underrepresented brands without overfitting.

Building on these prior works, our approach focuses exclusively on surveillance images, leveraging ResNet-18 with Focal Loss and selective augmentation to enhance classification performance in real-world settings.

III. Processing Pipeline

Our approach to vehicle manufacturer classification from surveillance images consists of several key processing stages. This section provides a high-level overview of the pipeline, while the subsequent sections detail specific aspects such as feature extraction and learning strategies.

1) Dataset Preparation: We utilize the CompCars dataset, specifically its surveillance subset, which includes images captured under real-world conditions.
2) Data Preprocessing: The dataset is cleaned and annotated, with image paths mapped to corresponding car manufacturers. We apply label encoding to convert textual class labels into numerical representations.
3) Data Augmentation: To address class imbalance, targeted augmentation techniques such as random flipping, rotation, and color jittering are applied selectively to underrepresented classes.
4) Model Selection and Training: We use ResNet-18, a well-established CNN architecture, and train it using Focal Loss to improve classification performance on rare car makes.

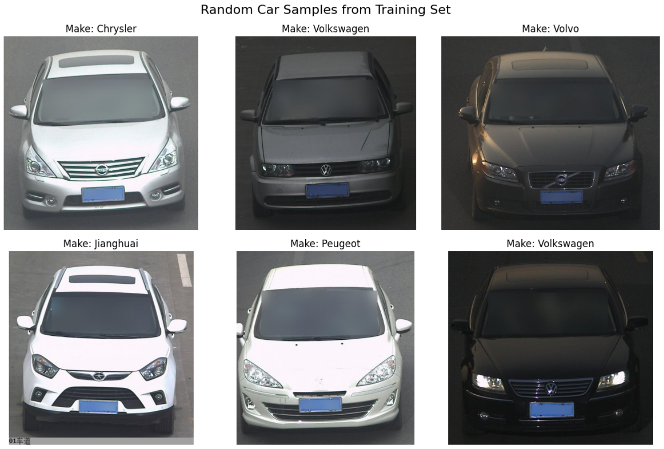

Fig. 1: Example of data samples

5) Evaluation: The trained model is evaluated using accuracy, loss, and confusion matrices, ensuring robust classification performance.

Each of these steps plays a crucial role in improving classification accuracy under real-world conditions. The following sections will elaborate on feature extraction, dataset partitioning, and learning strategies in greater detail.

IV. Signals and Features

This section describes the dataset, preprocessing steps, and data partitioning strategy used in our study. Since we focus on vehicle manufacturer classification from surveillance images, handling class imbalance, image variability, and data augmentation is critical for ensuring a robust model.

A. Dataset Overview

The dataset consists of 67 vehicle manufacturers, with images captured from traffic cameras at different angles.

Each image in the dataset is labeled with its respective vehicle manufacturer. However, the dataset exhibits a strong class imbalance, where certain manufacturers, such as Volkswagen, appear significantly more frequently than others, such as Infiniti. The most common manufacturer has 2526 training images, whereas the least represented has only 12. Without proper handling, this imbalance could lead to biased model predictions favoring overrepresented classes.

B. Data Preprocessing

To ensure consistency and improve training stability, several preprocessing steps are applied to all images before being used in the model:

- Resizing: Since the dataset consists of images with varying resolutions, all images are resized to 224×224 pixels to standardize input dimensions.
- Normalization: Images are converted to RGB format and normalized using the ImageNet mean and standard deviation to align with standard deep learning architectures.
- Data Augmentation: To counteract the effects of class imbalance and enhance model generalization, targeted

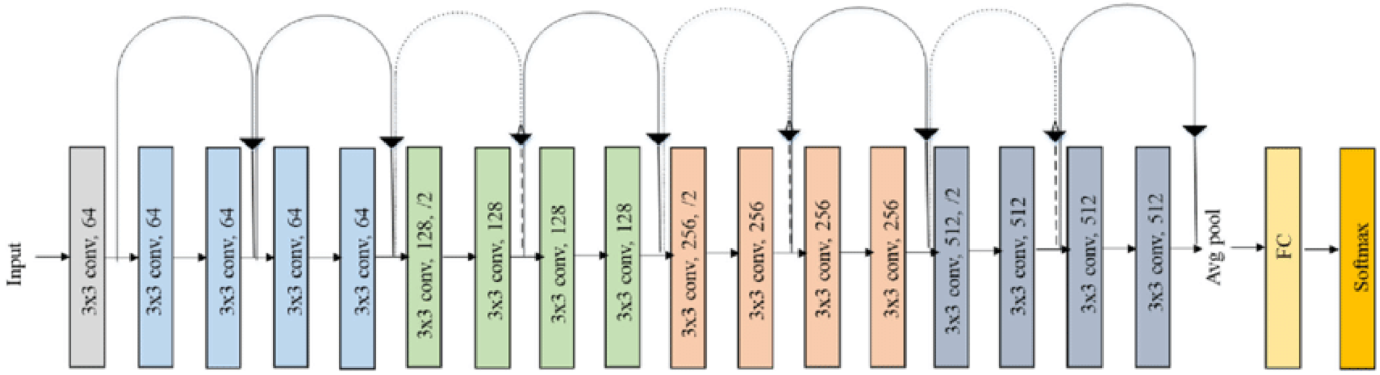

Fig. 2: ResNet-18 Architecture

augmentation is applied. This includes random horizontal flipping, brightness and contrast adjustments, rotation, affine transformations, and perspective distortion. Importantly, augmentation is selectively applied to underrepresented classes to prevent further bias toward well-represented brands.

C. Dataset Partitioning

The dataset is split into training, validation, and test sets following the predefined partitions provided by the dataset creators [1]. This ensures consistency with previous research and facilitates fair benchmarking.

- Training set: Used for model learning. Additional augmentation is applied to underrepresented classes.
- Validation set: Extracted from the training set before training begins, ensuring consistent evaluation across all epochs. This avoids data leakage and allows reliable hyperparameter tuning.
- Test set: Used for final model evaluation.

The dataset distribution reveals a significant imbalance, with a middle range of 300-500 images per class. To mitigate this, a threshold-based augmentation strategy is employed, where classes with fewer than 500 images in training receive additional synthetic data, while those with fewer than 100 images in validation and test sets are also augmented to maintain distribution consistency.

D. Example images

Fig. 1 presents sample images from the dataset, illustrating the challenges posed by surveillance-based vehicle classification, including occlusions, lighting variations, and viewpoint differences. These examples highlight the need for a robust preprocessing and augmentation pipeline to ensure reliable model performance.

V. Learning Framework

A. Model Architecture

For our task of car manufacturer classification, we employ ResNet-18, a deep convolutional neural network (CNN) known for its residual learning framework. ResNet-18 consists of 18 layers, primarily composed of convolutional blocks and identity mappings, allowing deeper networks to be trained effectively without the vanishing gradient problem [5]. The architecture is presented in the Fig. 2 is defined as follows:

- An initial convolutional layer with a 3×3 kernel, followed by Batch Normalization (BN) and ReLU activation.
- Four stages of residual blocks, where each block consists of two 3×3 convolutional layers and skip connections.
- A global average pooling (GAP) layer to reduce dimensionality.
- A fully connected (FC) layer for classification into $C = 67$ car brands.

The final output of the model is a vector $\boldsymbol{y} \in \mathbb{R}^C$, which is transformed into class probabilities using the *softmax activation function*:

$$P(y_i) = \frac{e^{y_i}}{\sum_{j=1}^{C} e^{y_j}}, \tag{1}$$

where $P(y_i)$ represents the probability of an image belonging to class i.

B. Loss Function: Focal Loss for Imbalanced Classes

A significant challenge in our dataset is class imbalance, where certain car manufacturers are overrepresented (e.g., Volkswagen: *2526 samples*) while others are underrepresented (e.g., Infiniti: *12 samples*). Traditional *Cross-Entropy Loss (CE)* tends to be biased toward majority classes, leading to poor generalization on minority classes.

To address this, we adopt *Focal Loss*, introduced in [6], which modifies the standard cross-entropy loss to focus more on hard-to-classify samples. The *Focal Loss* is formulated as:

$$FL(p_t) = -\alpha_t (1 - p_t)^\gamma \log(p_t), \tag{2}$$

where: - p_t is the softmax probability of the correct class, - α_t is a weighting factor for class balance, - γ is the focusing parameter (set to $\gamma = 2$) to down-weight well-classified examples.

By dynamically adjusting loss contribution based on confidence scores, Focal Loss ensures the model learns more effectively from difficult examples without being dominated by easy-to-classify majority class samples.

C. Optimization Strategy

We optimize the model using **Adam optimizer** with a learning rate of 10^{-4}. The network is trained for **30 epochs**, with early stopping triggered after **5 epochs** without validation improvement. Additionally, mini-batch size is set to **32**.

D. Implementation Details

The model is implemented in PyTorch, trained using Google Colab with a L4 GPU. Training takes approximately **10 minutes per epoch**, totaling 5 hours for full training.

VI. Results

The training process was conducted for a maximum of 30 epochs; however, early stopping was applied to prevent overfitting. The training stopped at **epoch 23**, as no significant improvement was observed in the validation loss for 5 consecutive epochs. This highlights the stability and effectiveness of the training process.

A. Training and Validation Performance

The learning process exhibited a rapid improvement in accuracy within the first few epochs, with the model reaching a validation accuracy of **89.66%** at the stopping point. The loss and accuracy curves at Fig. 3 and Fig. 4 demonstrate a steady decrease in both training and validation loss, suggesting that the network successfully learned meaningful representations without excessive overfitting.

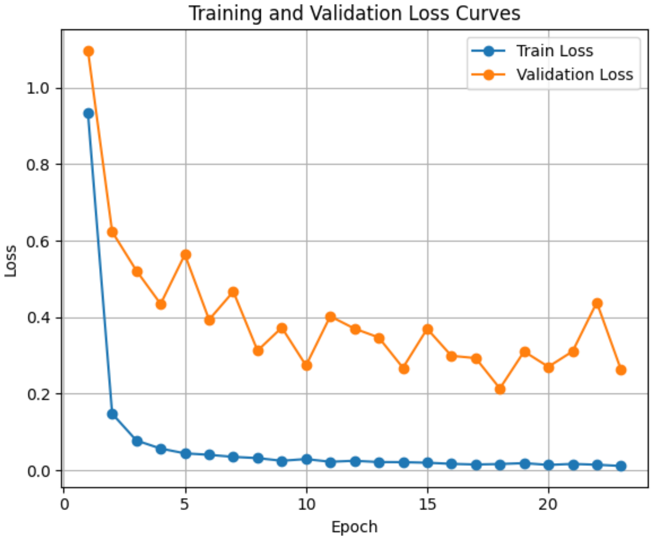

Fig. 3: Training and validation loss over epochs.

B. Test Set Evaluation

After training, the best model (saved at epoch 18) was evaluated on the test set, achieving a **test accuracy of 89.66%**. The **test loss was 0.2078**, indicating a well-generalized model.

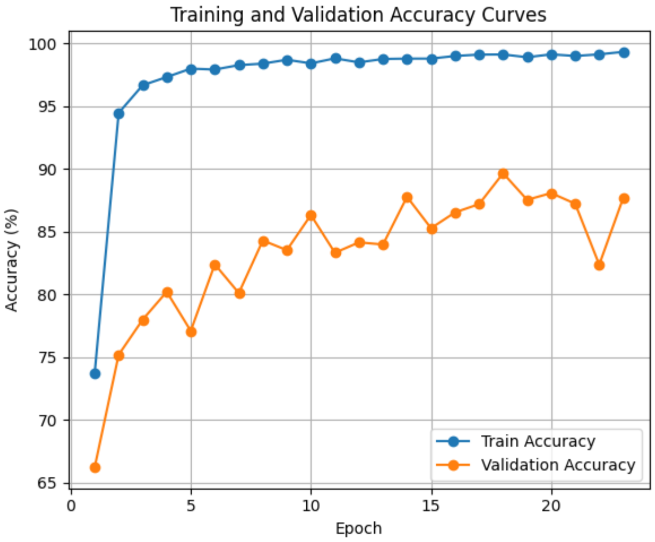

Fig. 4: Training and validation accuracy over epochs.

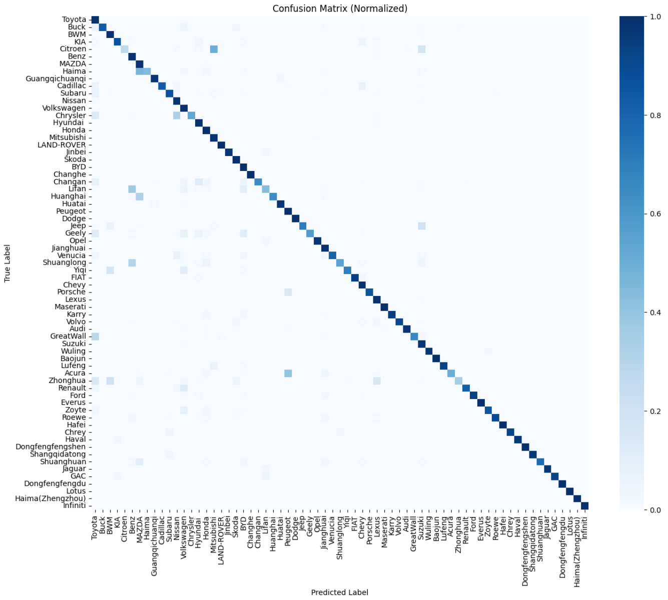

Fig. 5: Confusion matrix of the classification results.

C. Confusion Matrix Analysis

The confusion matrix at Fig. 5 provides insight into the classification performance across different car manufacturers. The majority of misclassifications occurred among brands with visually similar features (e.g., BMW and Mercedes-Benz).

TABLE 1: Performance Metrics on the Test Set

Metric	**Value**
Test Accuracy	89.66%
Test Loss	0.2078

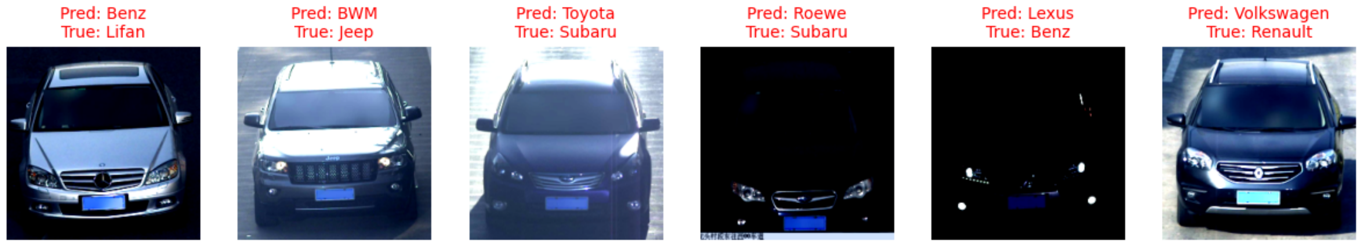

Fig. 6: Examples of misclassified images. The predicted and true labels are displayed below each image.

D. *Misclassified Examples*

To better understand the model's limitations, Fig. 6 displays six randomly selected misclassified images. These errors are often attributed to poor lighting conditions, occlusions, or visually similar car designs.

E. *Impact of Focal Loss and Data Augmentation*

Focal Loss played a crucial role in mitigating class imbalance by focusing more on hard-to-classify samples. Additionally, targeted data augmentation improved the model's ability to generalize, particularly for underrepresented classes. Without these techniques, preliminary experiments showed that accuracy dropped by approximately **4-5%**, demonstrating the necessity of these enhancements.

VII. Concluding Remarks

In this project, we tackle the task of car manufacturer classification using deep learning in the CompCars surveillance data set.

Our findings highlight the effectiveness of class-balancing strategies in improving classification performance for underrepresented categories. The combination of Focal Loss and selective augmentation proved beneficial, allowing the model to focus on hard-to-classify samples while avoiding overfitting to majority classes.

However, certain limitations remain. The model occasionally confuses similar-looking car brands, particularly those with shared design elements. Future work could explore fine-grained classification techniques, attention mechanisms, or self-supervised learning to further improve recognition accuracy. Additionally, experimenting with larger architectures such as ResNet-50 or Vision Transformers could yield further performance gains.

From an educational perspective, this project provided valuable hands-on experience in deep learning model training, dataset preprocessing, and evaluation techniques. One of the main challenges we encountered was dealing with imbalanced data, which requires careful selection of augmentation and loss functions to ensure fair learning across all classes. In addition, optimizing model efficiency for real-time applications remains an open area for further research.

In general, this study reinforces the importance of robust data pre-processing and thoughtful architecture selection when working with real-world image classification problems.

References

[1] L. Yang, P. Luo, C. Change Loy, and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3973–3981, 2015.

[2] Q. A. Al-Haija, M. A. Smadi, and S. Zein-Sabatto, "Multi-class weather classification using resnet-18 cnn for autonomous iot and cps applications," in *2020 International Conference on Computational Science and Computational Intelligence (CSCI)*, pp. 1586–1591, 2020.

[3] M. A. Hossain and M. S. A. Sajib, "Classification of image using convolutional neural network (cnn)," *Global Journal of Computer Science and Technology*, vol. 19, no. 2, pp. 13–14, 2019.

[4] M. Buzzelli and L. Segantin, "Revisiting the compcars dataset for hierarchical car classification: New annotations, experiments, and results," in *Sensors*, vol. 21, 2021.

[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

[6] Q. H. Chen, G., "Class-discriminative focal loss for extreme imbalanced multiclass object detection towards autonomous driving.," *Vis Comput*, vol. 38, pp. 1051–1063, 2022.