Detection and Classification of Vehicles using Deep Learning

Elmehdi Chenguiti, Elmir Zakaria, Othmane Himmi, Zie Abdoulaye Ouattara

K E Y W O R D S

Deep Learning
Deep Neural Network
Object Classification
Vehicle Classification
CNN
Quantization
Transfer Learning
Fine Tuning
Data Augmentation
Computer Vision
Image Processing
Traffic Surveillance
Image recognition
Traffic management
Xception
ConvNeXtBase
EfficientNetV2M
ResNet50

A B S T R A C T

This research paper delves into the realm of efficient vehicle classification in resource-constrained environments through a strategic combination of transfer learning, fine-tuning, and model quantization. Leveraging four distinct models—Xception, ConvNeXtBaseBase, EfficientNetV2M, and ResNet50—our study focuses on a meticulously curated dataset featuring 4800 tiny, low-resolution vehicle images. The dataset encompasses five categories: Bike, Car, Minibus, Pickup, and Truck, each with 800 images sized at 100 × 100 pixels and 96 dots per inch resolution. The deliberate choice of this compact dataset aligns with the challenges posed by limited labeled data and computational resources in real-world scenarios. Beyond achieving superior classification accuracy, our objective extends to deploying efficient models in low-resource settings. Through an in-depth analysis of the fine-tuned models, we identify the top-performing candidate. Subsequently, model quantization techniques are employed to compress the chosen model, ensuring its deployability in environments marked by computational constraints. This research not only advances the field of efficient vehicle classification but also offers insights into the practicality of deploying deep learning models in real-world applications constrained by limited resources. The amalgamation of transfer learning, fine-tuning, and model quantization serves as a comprehensive strategy to strike a balance between accuracy and efficiency in image classification tasks, particularly within vehicular contexts characterized by resource limitations.

1. Introduction:

In recent years, advancements in computer vision and deep learning have revolutionized image classification tasks, offering unprecedented accuracy in diverse applications. Transfer learning, a technique that leverages pre-trained models on large datasets, has become a pivotal approach for achieving state-of-the-art performance, especially when confronted with limited labeled data. In this research paper, we present our endeavors in employing transfer learning and fine-tuning strategies on four distinct models—Xception, ConvNeXtBaseBase, EfficientNetV2M, and ResNet50—applied to a unique dataset comprising 4800 tiny and low-resolution vehicle images. The dataset is meticulously curated, focusing on vehicles classified into five distinct categories: Bike, Car, Minibus, Pickup, and Truck. Each class consists of 800 vehicle images, each sized at 100 × 100 pixels with a resolution of 96 dots per inch (dpi). The deliberate choice of a smaller image size and resolution aligns with scenarios encountered in resource-constrained environments, where computational resources and storage capacity may be limited. Our objective extends beyond achieving superior classification accuracy; we address the need for deploying efficient models in low-resource settings. To this end, we conduct an in-depth analysis of the four fine-tuned models and identify the top-performing candidate. Subsequently, we employ model quantization techniques to compress the chosen model, ensuring that it remains deployable in environments characterized by restricted computational capabilities. This research contributes not only to the domain of efficient vehicle classification but also provides valuable insights into the feasibility of deploying deep learning models in real-world applications where resource constraints are a critical consideration. The amalgamation of transfer learning, fine-tuning, and model quantization serves as a comprehensive approach to strike a balance between accuracy and efficiency in image classification tasks, particularly within the context of vehicular scenarios with limited resources.

2. Related Work:

In response to the challenges outlined, recent research efforts have explored the application of deep learning, particularly Convolutional Neural Networks (CNN), in the domain of vehicle detection and classification [1–6]. In [1], a CNN-based system is proposed for real-time vehicle detection and classification using a low-quality monitoring camera. The study evaluates the performance under real-time constraints, comparing execution times on both CPU and GPU. Another approach, detailed in [2], employs the Faster R-CNN architecture to detect and classify distant vehicles in real-time scenarios, assessing performance across varied weather conditions. Addressing the issue of low resolution in classifying tiny objects, [3] introduces a method utilizing a generative adversarial network (GAN) with two CNNs.

This approach generates high-resolution images from low-resolution counterparts, enhancing classifier accuracy. Despite these contributions, it is noteworthy that limited research exists on deep learning-based vehicle classification for low-quality images collected by a distant, wide-angle surveillance camera designed for security purposes rather than traffic monitoring [1–6].

3. Vehicle dataset:

The Vehicle dataset presented in this study comprises 4,000 images, organized into five distinct categories: Bike, Car, Minibus, Pickup, and Truck, with each category containing 800 images. The images are standardized to a resolution of 100 × 100 pixels, maintaining a density of 96 dots per inch. This deliberate curation aligns with the practical challenges encountered in real-world applications, where limited labeled data and computational resources are common constraints. To address these challenges, the dataset has undergone data augmentation techniques, ensuring a diverse and enriched training set. The compact nature of this dataset serves as a valuable resource for researchers and practitioners, offering a realistic testbed for the development and evaluation of robust vehicle classification models under constrained conditions.

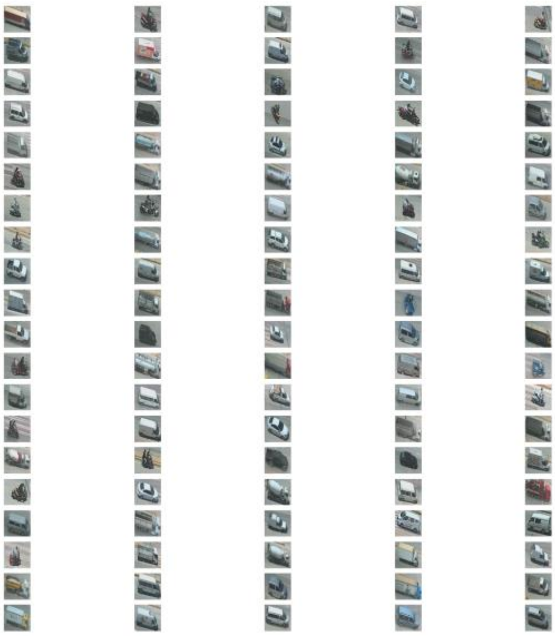

Figure 1 : Vehicle dataset

Figure 2 : Bike Category

Figure 3 : Truck Category

Figure 4 : Pickup Category

Figure 5 : Car Category

Figure 6 : Minibus Category

4. Methodology:

In the Methods section, we systematically refine our deep learning models for efficient vehicle classification, employing a comprehensive strategy. Initially, we harness the capabilities of transfer learning by employing pretrained models, namely Xception, ConvNeXtBase, EfficientNetV2M, and ResNet50, all pretrained on the ImageNet dataset. Throughout the process, we strategically freeze all weights of the models, excluding the last fully connected layers that we incorporate for our specific vehicle classification task. After an initial training phase where we observe the stability of both validation and training sets, signifying convergence, we proceed to unfreeze the weights of the entire model. Subsequently, we fine-tune the now unfrozen models for additional epochs, utilizing a notably reduced learning rate. This meticulous approach aims to optimize the adaptability of Xception, ConvNeXtBase, EfficientNetV2M, and ResNet50 to our curated vehicle dataset, ensuring superior classification accuracy while accommodating the challenges posed by limited labeled data and computational resources in real-world scenarios.

4.2. Transfer Learning:

Transfer learning consists of taking features learned on one problem, and leveraging them on a new, similar problem. For instance, features from a model that has learned to identify racoons may be useful to kick-start a model meant to identify tanukis.

Transfer learning is usually done for tasks where your dataset has too little data to train a full-scale model from scratch.

The most common incarnation of transfer learning in the context of deep learning is the following workflow:

Firstly, Take layers from a previously trained model.

Secondly Freeze them, to avoid destroying any of the information they contain during future training rounds.

Thirdly Add some new, trainable layers on top of the frozen layers. They will learn to turn the old features into predictions on a new dataset.

Finally Train the new layers on your dataset.

4.3. Fine Tuning:

consists of unfreezing the entire model you obtained above (or part of it), and re-training it on the new data with a very low learning rate. This can potentially achieve meaningful improvements, by incrementally adapting the pretrained features to the new data.

4.4. Quantization:

Quantization in deep learning is the process of reducing the precision of weights and activations in a neural network, typically using fewer bits to represent them. This helps reduce memory usage and computational resources during inference. Weight quantization focuses on reducing the precision of network parameters, while activation quantization targets intermediate values. Binary quantization restricts values to binary (-1 or 1). Though quantization may lead to some loss in model accuracy, techniques like fine-tuning aim to mitigate this impact.

```
Quantized Custom ResNet's Accuracy on Testing Set: 95.20%
```

4.5. Xception:

4.5.1. Description:

Xception, short for "Extreme Inception," is a deep neural network architecture that introduces a novel approach to feature extraction using depthwise separable convolutions. Developed by Google, it aims to capture complex patterns efficiently while reducing the number of parameters.

4.5.2. Structure:

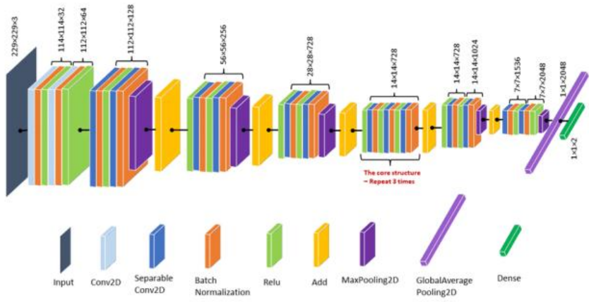

Figure 7 : Xception Stucture

4.6. ResNet50:

4.6.1. Description:

ResNet50, short for "Residual Network with 50 layers," is a variant of the ResNet architecture. ResNet introduced the concept of residual learning, using skip connections to ease the training of very deep networks. ResNet50, specifically, consists of 50 layers and is widely used for various computer vision tasks.

4.6.2. Structure

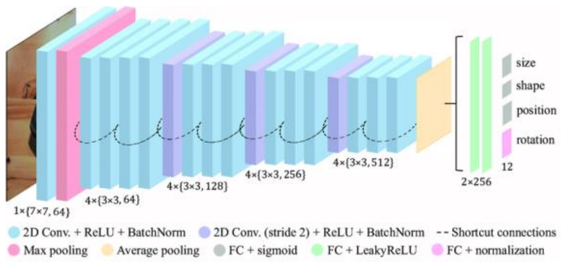

Figure 8 : ResNet50 Stucture

4.7. EfficientNetV2M:

4.7.1. Description:

EfficientNetV2M is part of the EfficientNet family of models known for achieving high accuracy with fewer parameters. Developed to address resource constraints, EfficientNetV2M specifically focuses on efficient model scaling, balancing model size and performance for mobile and edge computing applications.

4.7.2. Structure:

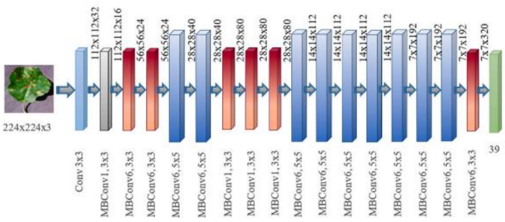

Figure 9 : EfficientNetV2M Structure

4.8. ConvNeXtBase:

4.8.1. Description

ConvNeXtBase is a convolutional neural network architecture designed to enhance feature representation by incorporating cross-stage hierarchical connections. By introducing cross-stage connections, ConvNeXtBase facilitates better information flow between different stages of the network, improving model performance.

4.8.2. Structure

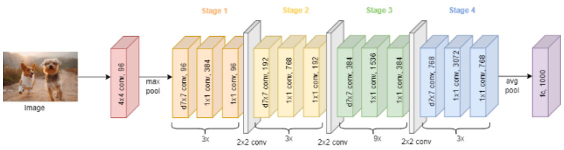

Figure 10 : ConvNeXt Structure

5. Results:

5.1. Xception

In the subsequent section, we present the outcomes of training the **Xception** model, examining the variations in training and validation loss, alongside comprehensive insights from confusion matrices and classification reports.

5.1.1. **Xception** Learning Curve**:**

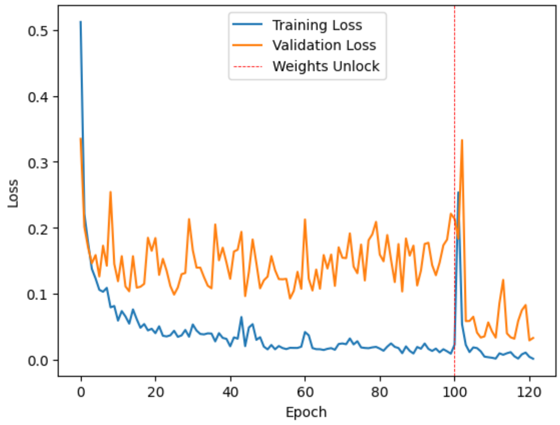

Figure 11 : Xception Training/Validation Loss Evolution

5.1.2.Custom **Xception's** Confusion Matrix:

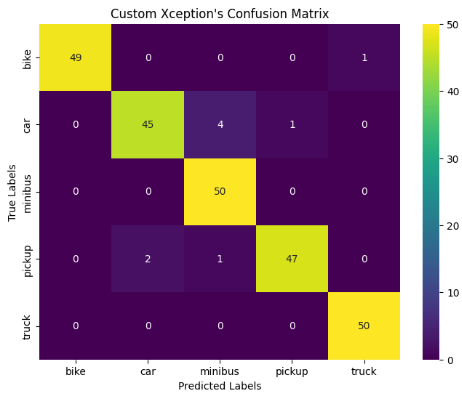

Figure 12 : Custom Xception's Confusion Matrix

5.1.3. Custom **Xception's** Classification Report:

Custom Xception Classification Report:

	precision	recall	f1-score	support
bike	1.00	0.98	0.99	50
car	0.96	0.90	0.93	50
minibus	0.91	1.00	0.95	50
pickup	0.98	0.94	0.96	50
truck	0.98	1.00	0.99	50
accuracy			0.96	250
macro avg	0.97	0.96	0.96	250
weighted avg	0.97	0.96	0.96	250

Figure 13: Xception's Report

5.2. **ResNet50**:

In the subsequent section, we present the outcomes of training the **ResNet50** model, examining the variations in training and validation loss, alongside comprehensive insights from confusion matrices and classification reports.

5.2.1 . **ResNet50** Learning Curve:

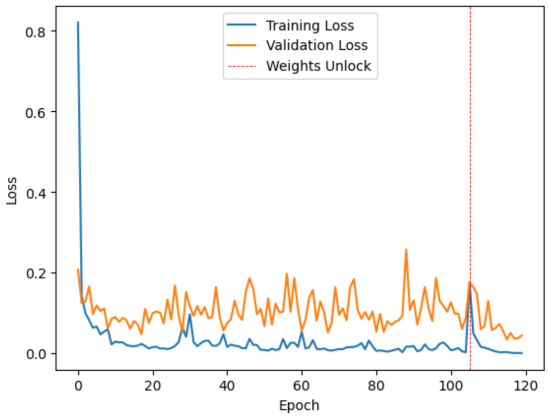

Figure 14 : ResNet50 Training/Validation Loss Evolution

5.2.2 . Custom **ResNet50** Confusion Matrix:

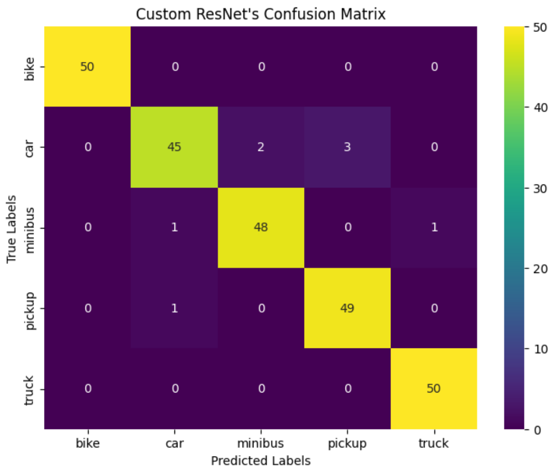

Figure 15 : Custom ResNet50 Confusion Matrix

5.2.3 . Custom **ResNet50** Classification Report:

Custom ResNet Classification Report:

	precision	recall	f1-score	support
bike	1.00	1.00	1.00	50
car	0.96	0.90	0.93	50
minibus	0.96	0.96	0.96	50
pickup	0.94	0.98	0.96	50
truck	0.98	1.00	0.99	50
accuracy			0.97	250
macro avg	0.97	0.97	0.97	250
weighted avg	0.97	0.97	0.97	250

Figure 16 : Custom ResNet50 Classification Report

5.3. EfficientNetV2M:

In the subsequent section, we present the outcomes of training the EfficientNetV2M model, examining the variations in training and validation loss, alongside comprehensive insights from confusion matrices and classification reports.

5.3.1. **EfficientNetV2M** Learning Curve:

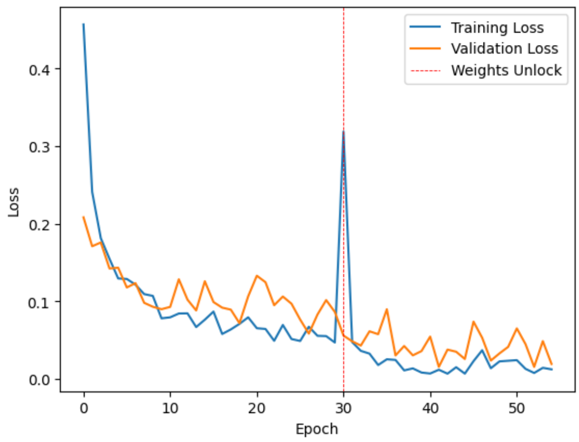

Figure 17: EfficientNetV2M Training/Validation Loss Evolution

5.3.2. Custom **EfficientNetV2M** Confusion Matrix:

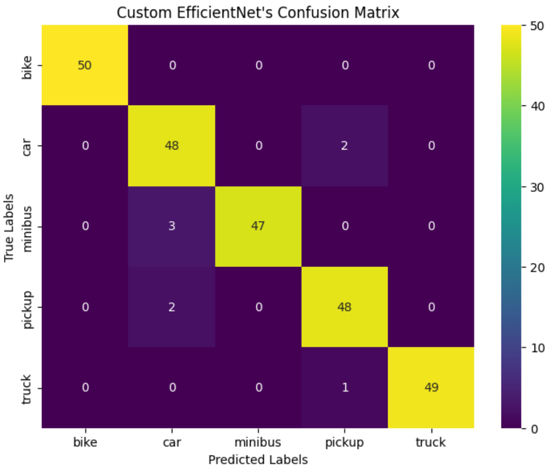

Figure 18: Custom EfficientNetV2M Confusion Matrix

5.3.3. Custom EfficientNetV2M Classification Report:

Custom EfficientNet Classification Report:

	precision	recall	f1-score	support
bike	1.00	1.00	1.00	50
car	0.91	0.96	0.93	50
minibus	1.00	0.94	0.97	50
pickup	0.94	0.96	0.95	50
truck	1.00	0.98	0.99	50
accuracy			0.97	250
macro avg	0.97	0.97	0.97	250
weighted avg	0.97	0.97	0.97	250

Figure 19: Custom EfficientNetV2M Classification Report

5.4. ConvNeXtBase:

In the subsequent section, we present the outcomes of training the ConvNeXtBase model, examining the variations in training and validation loss, alongside comprehensive insights from confusion matrices and classification reports.

5.4.1. **ConvNeXtBase** Learning Curve:

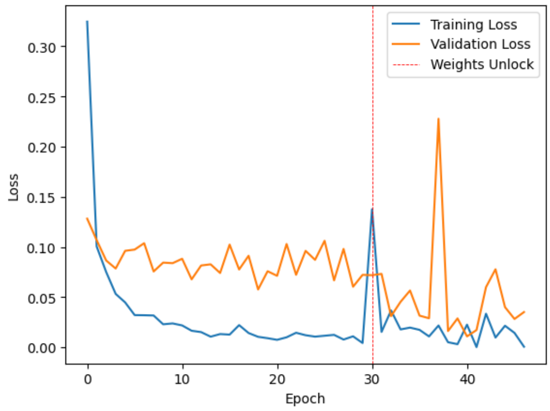

Figure 20 : ConvNext Training/Validation Loss Evolution

5.4.2. Custom ConvNeXtBase Confusion Matrix:

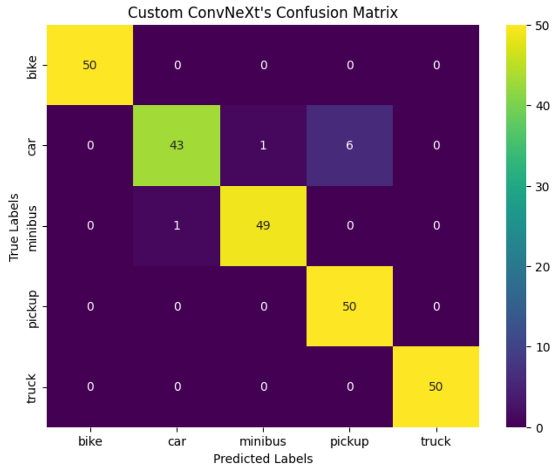

Figure 21 : Custom ConvNext Confusion Matrix

5.4.3. Custom ConvNeXtBase Classification Report:

Custom ConvNeXt Classification Report:

	precision	recall	f1-score	support
bike	1.00	1.00	1.00	50
car	0.98	0.86	0.91	50
minibus	0.98	0.98	0.98	50
pickup	0.89	1.00	0.94	50
truck	1.00	1.00	1.00	50
accuracy			0.97	250
macro avg	0.97	0.97	0.97	250
weighted avg	0.97	0.97	0.97	250

Figure 22: Custom ConvNeXtBase Classification Report

6. Discussion:

In this section, we conduct a comparative analysis of four prominent models—ResNet50, Xception, ConvNeXtBase, and EfficientNetV2M—utilized in our vehicle classification task. The accuracy metrics on the testing set reveal comparable performance across custom ResNet50 (96.80%), custom Xception (96.40%), custom ConvNeXtBase (96.80%), and custom EfficientNetV2M (96.80%). Despite the apparent similarity in accuracy, our discussion extends beyond numerical values to consider model complexity and resource efficiency.

ResNet50 emerges as the optimal choice among the three models with an accuracy of 96.80%, primarily due to its superior efficiency in terms of parameterization. Although its accuracy aligns with that of ConvNeXtBase and EfficientNetV2M, ResNet50 achieves this with the lowest number of parameters. The emphasis on model efficiency becomes crucial in real-world scenarios marked by limited computational resources, making ResNet50 the pragmatic choice for deployment in resource-constrained environments.

While both ConvNeXtBase and EfficientNet demonstrate commendable accuracy, their higher parameter count could impede adaptability in scenarios with limited computational resources.

In conclusion, the balance between accuracy and efficiency positions ResNet50 as the most suitable candidate for our vehicle classification task. Its streamlined architecture with fewer parameters ensures a judicious use of computational resources without compromising on classification precision. This strategic choice aligns with the overarching goal of deploying efficient models in real-world applications characterized by limitations in both labeled data and computational capacity.

Model	Size (MB)	Number of Parameters
Xception	88	22.9M
ResNet50	98	25.6M
EfficientNetV2M	220	54.4M
ConvNeXtBase	338.58	88.5M

Figure 23 : Models size and N° of params

7. Conclusion:

In concluding this research endeavor focused on efficient vehicle classification through deep learning, we emphasize the importance of model selection, fine-tuning strategies, and the potential adoption of early exit. Our comparative analysis identifies ResNet50 as the optimal choice, balancing accuracy with streamlined parameter efficiency. This intentional selection ensures the deployability of models in resource-constrained environments, addressing real-world challenges where computational resources and labeled data are limited. While the current project did not incorporate early exit, it stands as a promising direction for future exploration. The implementation of early exit, a technique designed to enhance a model's computational efficiency, could further refine our ResNet50-based model. By introducing exit points in the neural network and evaluating predictions at intermediate stages, early exit offers the potential to improve efficiency without compromising accuracy. As we conclude this research, the broader implications extend beyond the immediate context of vehicle classification. Our findings contribute to the evolving landscape of deep learning applications in resource-constrained scenarios. The strategic amalgamation of transfer learning, fine-tuning, quantization and the potential incorporation of early exit with ResNet50 provides a comprehensive framework for navigating challenges in real-world image classification tasks. As technology continues to advance, the application of these methodologies, particularly with ResNet50, remains adaptable and holds promise for continued refinement in future projects.

References:

[1] Vijayaraghavan, V., & Laavanya, M. (2019). Vehicle Classification and Detection using Deep Learning.(https://www.ijeat.org/wp-content/uploads/papers/v9i1s5/A10061291S52019.pdf)

[2] Jagannathan, P., Rajkumar, S., Frnda, J., Divakarachari, P. B., & Subramani, P. (2021). Moving Vehicle Detection and Classification Using Gaussian Mixture Model and Ensemble Deep Learning Technique. (https://www.researchgate.net/publication/351925609_Moving_Vehicle_Detection_and_Classification_Using_Gaussian_Mixture_Model_and_Ensemble_Deep_Learning_Technique)

[3] Song, H., Liang, H., Li, H., Dai, Z., & Yun, X. (2019). Vision-based vehicle detection and counting system using deep learning in highway scenes. (https://etrr.springeropen.com/articles/10.1186/s12544-019-0390-4)

[4] Chen, Y., & Li, Z. (2021). An Effective Approach of Vehicle Detection Using Deep Learning. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9356827/)

[5] LI, S., LIN, J., LI, G., BAI, T., WANG, H., & PANG, Y. (2018). Vehicle type detection based on deep learning in traffic scene. (https://www.researchgate.net/publication/325085672_Vehicle_type_detection_based_on_deep_learning_in_traffic_scene)

[6] Tas, S., Sari, O., Dalveren, Y., Pazar, S., Kara, A., & Derawi, M. (2022). Deep Learning-Based Vehicle Classification for Low-Quality Images. (https://www.mdpi.com/1424-8220/22/13/4740)

Acknowledgements

We would like to express our sincere appreciation to Professor Mehdi Zakroum for his invaluable guidance and unwavering support throughout our project. Professor Zakroum's expertise and dedication played a pivotal role in shaping the success of our endeavor.

His profound knowledge and insightful feedback significantly enriched our understanding of the subject matter, providing us with a strong foundation for tackling challenges and making informed decisions. Professor Zakroum's mentorship went beyond the classroom, as he generously shared his wealth of experience, inspiring us to explore innovative solutions and fostering a collaborative learning environment.

Our commitment to excellence and passion for the subject were fueled by his motivating influence, encouraging us to strive for higher standards in our research and development efforts. Professor Zakroum's approachable demeanor and willingness to invest time in addressing our queries greatly enhanced the overall learning experience.

In acknowledging Professor Mehdi Zakroum, we express our deepest gratitude for his instrumental role in our project's success, shaping not only our academic journey but also contributing significantly to our personal and professional growth. His impact will undoubtedly resonate as a lasting influence on our future endeavors.

Ecole Supérieure
d'Informatique et du Numérique
COLLEGE OF ENGINEERING & ARCHITECTURE