Vehicle classification under real-world conditions: A comparative analysis of CNN and EfficientNet

Jinyang Ren *

School of Computer Science, University of Leeds, Leeds, United Kingdom

* Corresponding Author Email: wpzb0774@leeds.ac.uk

Abstract. This paper uses deep learning techniques for vehicle type classification in real world traffic situations. A dataset of 300 images of cars, trucks, and buses was created. Images were resized to 224x224 pixel dimensions. For deep learning models, images were preprocessed by way of flipping, rotating, and changing brightness to diversify the dataset. For this study, we used compared two deep learning methods: A simple convolutional neural network, and in this case, EfficientNet , which involves transfer learning. During this study, the basic CNN was able to classify images with 68 % accuracy while EfficientNet was able to classify images with 96 % accuracy. This research strongly indicates, and it is to be expected, that the selection of advanced network designs aids in the classification of vehicles. More work needs to be done in optimizing the model's reliability and practicality, and this can be done by increasing the size of the dataset, adding complexity to the data, and testing with different network designs.

Keywords: Vehicle classification; convolutional neural network; EfficientNet; deep learning.

1. Introduction

In recent years, the rapid development of artificial intelligence and computer monitoring has greatly enhanced the processing power and ability to understand visual information of machines [1] . Object detection and image classification are core technologies in the field of computer vision, widely used in intelligent transportation systems [2]. The categorization of vehicles is relevant in real-life applications like traffic control, autonomous driving, and city traffic control [3]. As the society is evolving constantly, there is a constant need of smart transportation. Individuals wish that the transportation systems should be able to automatically perform tasks like vehicle recognition, traffic flow statistics and violation detecting. All these functions employ the accurate vehicle classification technology to examine the approach to enhancing the accuracy and stability of vehicle classification in practical settings, which possesses exceptional practical value. Despite the great advances achieved in image recognition with deep learning, the accurate and stable classification of vehicles in the real world and in harsh environments is still extremely challenging. Most of the models are very precise when the conditions are ideal in the laboratory, including good illumination, clear shooting angles, and coherent backgrounds. In real road environment, performance of recognition is usually influenced by numerous unsteady elements, including variations in lighting, obscuration by structures or other vehicles, shadow coverage, and complicated backgrounds. These may lead to the deterioration of the quality of feature information that the model extracts, leading to a decrease in classification accuracy and misjudgment [4]. In cases of vehicles that are partially blocked, the main features that the model can derive would be highly inadequate and it would thus be hard to make the right decisions. To address such issues, we have invested our efforts in the classification of vehicles in tricky real-life situations, studying how environmental factors (such as change of lighting, occlusion, and complicated background features) can affect the model and conduct study on viable ways to improve the model. Through the analysis of the sample of errors in the model, we have given a summary of the error patterns and explained the inadequacies of the current methods. We applied convolutional neural networks to train and test self collected vehicle dataset in the experimental part. To ensure the experimental setup is more similar to the physical road scenario, at the stage of data augmentation, we implemented brightness adjustment, shadow occlusion, and partial blurring to assess the model performance in terms of accuracy, precision and recall [5]. Based on the available literature, to investigate the experiment in more depth, we also had discussions on some of the optimization schemes, such as more effective data augmentation schemes, better preprocessing programs, and adjusting the appropriate parameters within a reasonable range. This paper presents a point of reference in the creation of viable vehicle classification systems in real world situations with the comparison of various experiments.

2. Methods

2.1. Dataset Preparation

In this study , we used a vehicle image dataset that we collected ourselves for model training and validation. The dataset consists of nearly 300 RGB images in the categories of cars, buses, and trucks. These images were resized to the same pixel dimensions in accordance with the requirements of experimental accuracy and model input size [6]. We chose to split the dataset into training and validation sets at an 8:2 ratio. The department aims to assess the model’s recognition capability on new data to test how well the model can generalize. During image preprocessing, pixel values were normalized within a specified range to speed up the convergence of training and to improve the training stability. We applied several data augmentation methods to increase the diversity of the training dataset, such as random rotations, flips, and adjustments to the image brightness. These methods help the model in real-world traffic scenarios where the lighting, angle, and occlusion of the images vary. Figure 1. shows a sample of the images in the dataset to illustrate the variations in photography conditions and vehicle categories[7] .

Figure 1. Example image of the dataset.

2.2. Models Based on Convolutional Neural Networks

Convolutional neural networks are used as one of the primary comparison models in this experiment. Convolutional neural networks (CNNs) are able to automatically find valuable spatial features in images. This contains edges, contours and textures which makes them the most popular structure in image classification tasks [8]. Convolutional neural networks operate by finding the representation of all the features through convolutional layers, the dimensionality of the features is diminished by pooling layers and the most significant information is retained through the fully connected layers before the final classification task can be performed by reflecting the features to every category. In the same experimental setting, we constructed two further networks to carry out comparative experiments [9]. We use simple convolutional network structure as the baseline model that consists of several layers of convolution, pooling, and a single fully connected layer. It is less complex and simpler to train, but in some sense unimpressive, because it cannot extract features, and is thus difficult to achieve high accuracy in complex cases.

2.3. Hyperparameter Configuration

For the duration of the training process, we set hyperparameter values based on assessed model performance and stability. With an appropriately set learning rate of 0.001, model weight updates are controlled so stability is maintained during training. The Adam optimizer has an adaptive learning rate, and is one of the better performing optimizer for image recognition [10]. For multi-class classification, the classification cross-entropy is applicable. In the evaluation process, we concentrated on determining the accuracy, but also recorded the loss to quantify the success of the training. The two models were trained for 20 and 30 epochs respectively. This amount of training ensured that the models were sufficiently trained and the training did not result in overfitting.

3. Results and Discussion

3.1. Experimental Results

This experiment contrasts the basic convolutional neural network model with the EfficientNet model. Bar charts in Figure 2 show the accuracy of the classifications of the two models. The horizontal axis is the model type, while the vertical shows the accuracy value. The experimental results show an obvious stark contrast between the two models. The accuracy value of plain convolutional neural network is 68%, while the efficient network achieves an accuracy value of 96%. This huge contrast shows the model's structure and the use of transfer learning are crucial in determining the final classification [11].

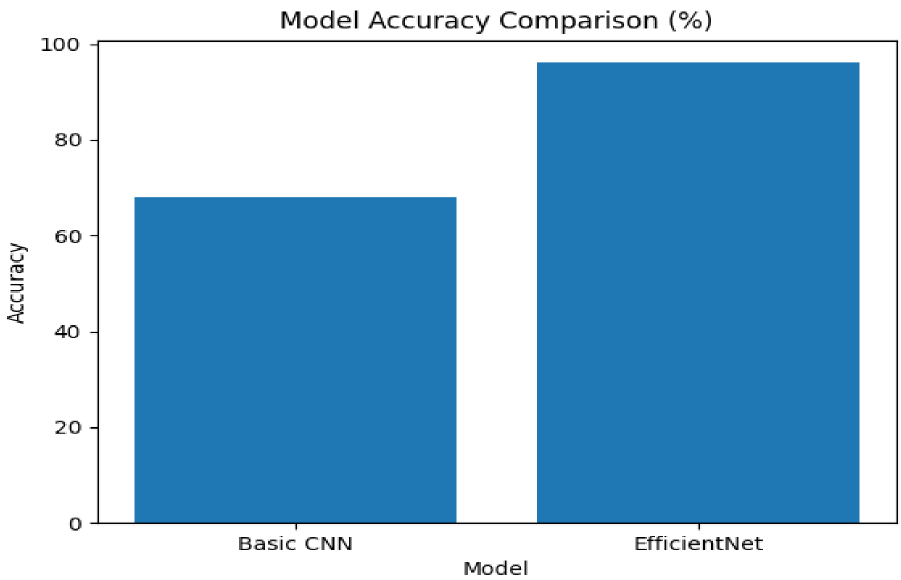

Figure 2. Comparison of the two types of networks.

3.2. Results Analysis

The structure and complexity of the network directly correlate to the classification results. According to the results of the experiment, the efficient network model leveraging transfer learning outperforms basic convolutional neural network models . The transfer learning models offer pre-trained weights, and they are a better structure, meaning that a model can detect subtler features. The effect is that the model can still recognize when the task is complex. Basic convolutional neural networks struggle to exhibit robust features, and they are quickly made irrelevant by the environment, meaning they can injury fix themselves. We noticed that the transfer learning models use learned knowledge more readily than their predecessors. Small datasets require a model to think, but shift and address goals. The point is that pre-trained weights help adapt models to a surprise.

3.3. Discussion

The efficient network achieved a prominent rank in this trial, yet many options exist for enhancements. The first factor which can formulate improvements to this model is the limited data. The dataset can be improved by data harvesting and including incidence with varying conditions of light, degrees of occlusion, and diverse background, which can augment the uniformity and generalize capability of the model, and this is mostly due to the small dataset as it severely restricts the differences in the augmentation parameters. Second, the augmentation techniques in this trial were outdated. There are many techniques which are unexplored, and which include random cropping, color dithering and noise injection, and are posited on the models to maintain stability in the presence of various perturbations, which is especially important for the goal of this direct model. The other improvements the model can receive are developed optimizations of the pretrained models which still exist due to the blunt adjustable techniques implemented in the upper most layers of the network. The number of layers adjustable must be optimized in order to improve the adaptability and data distribution alignment without crossing the threshold into overfitting. The last major improvements are reiterated model comparisons in conjunction with different networks, in order to expand upon MobileNet and ResNet and include supplementary models optimal for a classification framework [12].

4. Conclusion

This paper compares the performance of simple convolutional neural networks (CNNs) and efficient networks in real-world vehicle classification. The results show that the efficient network based on transfer learning has a significant advantage, achieving an accuracy of 96%, far exceeding the 68% of simple CNNs.Therefore , we believe that appropriately selecting the model structure and applying transfer learning can effectively improve the accuracy and stability of vehicle classification in complex environments. In future work, we will further improve the model training effect by expanding the dataset, refining data augmentation, and fine-tuning strategies .Some of the viewpoints and research presented in this paper have practical value and can provide a reference for the design of vehicle classification systems in intelligent transportation. With continuous optimization of models and data, vehicle classification technology will be more widely applied in real-world traffic scenarios.

Reference

[1] Yoshua Bengio et al. (2015) Deep Learning p.436–444 Link: https: // www.nature.com/articles/nature14539.

[2] Yun Hui Lin, Qingyun Tian (2020) Deep Learning in Intelligent Transportation Systems Section 1 / Introduction Link: https: //arxiv.org/abs/2002.09844.

[3] Benjamin Coifman a, Hunter Danford b, Mo Wu c, Yuyi Chang d, Keith Redmill d (2026) Vehicle Detection and Classification Section 1-2 Link: https :// www. sciencedirect.com/science/ article/pii/ S0968090X17300917.

[4] Alex Krizhevsky (2017) ImageNet Classification with Deep CNN p.1097-1105 Link: https://dl.acm.org/doi/10.1145/3065386.

[5] Richard Szeliski (2022) Computer Vision: Algorithms and Applications Section2-3 Link: https://szeliski.org/Book/.

[6] Alex Krizhevsky et al. (2017) ImageNet Classification with Deep Convolutional Neural Networks p.1097-1105 Link: https://dl.acm.org/doi/10.1145/3065386.

[7] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, Youngjoon Yoo (2019) A Survey on Image Data Augmentation Section2-3 Link: https://arxiv.org/abs/1905.04899.

[8] Yann LeCun (1998) Gradient-Based Learning Applied to Document Recognition p.2278-2324 Link: http://yann.lecun.com/exdb/publis/pdf/lecun-98.pdf.

[9] Mingxing Tan (2019) EfficientNet: Rethinking Model Scaling for CNNs Section3-4 Link: https://arxiv.org/abs/1905.11946.

[10] Diederik P. Kingma (2014) Adam: A Method for Stochastic Optimization Section 2 Link: https://arxiv.org/abs/1412.6980.

[11] Adam Falkowski, Giovanni Grilli di Cortona, Zahra Tabrizi (2018) Deep Learning for Visual Recognition: A Review Section3 Link: https://arxiv.org/abs/1802.08296.

[12] Karen Simonyan (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition (VGG) Section4 Link: https://arxiv.org/abs/1409.1556.