An Optimal BDCNN ML Architecture for Car Make Model Prediction

Kritika Kashyap, Rohit Miri

Dr. CV Raman University, Bilaspur (C.G.), India

Corresponding Author: Kritika Kashyap, Email: kritikaphd2022@gmail.com

Abstract: We take on the challenge of classifying car photos, from the most general car type to the precise make, model, and year of the vehicle for a given input. Analyzing pre-existing datasets, we find that the CompCars-SV are a great place to begin our classification project. We demonstrate that convolutional neural networks can obtain a classification accuracy of more than 90% on the most difficult task. Due to a skewed mix between training and testing, this impressive result isn't really typical of how people do in the actual world. Using an ML system for car detection, we automatically generate a vehicle-tight bounding box for each picture, which we disseminate to the full dataset together with the existing (but limited) type-level annotation. We have designed and implemented car classification algorithms to analyze this car dataset, two of which take advantage of the hierarchical nature of car annotations. According to our research, a more precise classification of car type at a finer resolution now achieves an accuracy of 99.25%. It serves as a baseline benchmark for future research. Focusing on "vehicle" tasks, this work intends to bring attention to the vision community's lack of attention to these tasks compared to other objects. The important reason getting higher accuracy is extraction of binary descriptor (BD) feature using edge detection before training the CNN. This step reduced the size of the car dataset; hence network took less time to get trained. From the result outcomes shown it is clear that the presented network architecture having 31 layers of 2d convolutional layer, batch normalization, maxpool, ReLU, fully connected layer and Softmax classifier layer, has given higher accuracy. Numerous relevant car-related issues and solutions have yet to be carefully examined and researched, according to our findings. Car model categorization, model verification, and attribute prognosis are just a few examples of how the dataset might be put to use.

Keywords: BDCNN; low complex layer graph; CompCar-SV; edge

1. INTRODUCTION

With the invention of the automobile, people were able to travel more easily and quickly from one location to another. Cars have become an essential method of transportation in our modern society. Having a car is often seen as a way to represent one's socioeconomic position or as a barometer of one's own socioeconomic standing. According to many automotive aficionados, the automobile has evolved into a topic of interest. Overall, the demand for automobiles
There are numerous applications for automated car model analysis in an intelligent transportation system, including regulation, classification, and indexing. Automating and speeding up the payment of tolls from the lanes, for example, can be done cheaply and easily using fine-grained car categorization, depending on different prices for different sorts of cars. The look of an automobile can be verified in video surveillance applications to help trace a vehicle over a network of several cameras when car plate recognition fails. Car verification algorithms can be used in post-event investigations to locate similar vehicles in the database. The examination of automobile models has a considerable impact on personal vehicle use. People often look at other cars on the road while they are considering purchasing a vehicle. It's easy to imagine a smartphone app that can show a user all the information about an automobile when they take a picture of it. An app like this will make it much easier for consumers to find out about a vehicle that they don't recognise. In other applications, producers and buyers may benefit from the recommendation of similar-looking vehicles based on a car's appearance and its predicted popularity.

Car model analysis receives little attention from the computer vision community, despite the fact that it is a hot topic for both academics and practitioners. It is our opinion that the scarcity of high-quality datasets in this field is limiting the investigation of the community. Therefore, we've compiled and organised what we term "Comprehensive Cars," abbreviated simply to "CompCars," a massive and comprehensive image repository. The "CompCars" dataset contains 44,481 photographs of 281 automobile models from surveillance-nature. It is now possible to test the performance of a wide range of computer vision algorithms on the new dataset. Real-world applications and unique research subjects can also benefit from this technology. Additionally, the dataset's multi-scenario
structure allows for cross-modality study. Section 3 provides a full explanation of CompCars.

We demonstrate various intriguing applications using the dataset, such as automobile model classification and testing based on convolutional neural network (CNN) [12], to prove its utility and urge the community to investigate new unique research ideas.

The paper is organized into six sections. Section 2 discusses the related work on vehicle recognition. Section 3 discusses about the characteristics of the CompCars dataset of cars details and brief about the make and models covered. Section 4 discussed about the applications of dataset in developing machine learning models of cars or vehicle recognition. Section 5 covers the research methodology of this research work to get better accuracy of recognition. The methodology covering the execution step details. Section 6 shows the results and its discussion in terms of accuracy and losses and its comparison with previous algorithms and the last section is about the conclusion of the manuscript.

2. Related Work
In [11] Y.Ren et al. discussed about “many smart applications use vehicle-type categorisation (VTC), license-plate recognition (LPR), and make and model registration (MMR) for vehicle analysis. MMR supports LPR in these tasks. We present a unique framework to identify moving vehicles and MMR using convolutional neural networks in this study. First, convolutional neural networks were trained and tested on car frontal pictures. Our vehicle MMR identification accuracy is 98.7% using our suggested framework.” In [12] F. Tafazzoli et al. shown “Vehicle Make and Model Recognition (VMMR) supports ITS/AVS. Real-time, accurate VMMR saves resources. VMMR demands multiclass classification with inter- and intra-make ambiguity. A 9,170-car-category picture dataset1 improves manufacturer and model identification. Low-light, partial, or occluded VMMR photographs. Extensive baseline trials enhanced outcomes. These strategies boost real-world VMMR”. In [13] author discussed about “Vehicle-based VMMR. This job demands expertise. Aesthetically, automotive frontal structural components differ. Structures have traits. Car logos. Intra-brand models enable brand class subclasses. Similar courses. Step 2 trains subclass-specific discriminant subregion classifiers. Patches improved this article's big car photo collection”. In [14] study about highlights car-related visual tasks, which the vision community has ignored. We demonstrate that many fascinating car-related problems and applications remain untouched. This work describes our ongoing collection of "CompCars" a vast dataset of car photos, interior and exterior components, and extensive features for automotive research. Cross-modal surveillance and web-nature datasets. Datasets categorise, verify, and forecast automotive models. We discuss automobile issues and applications. In [15] Linjie Yang et al. shows Computer-identified cars Traffic control, video monitoring, and identification need eyesight. Computers struggle with vehicle photo classification due to fine-grained properties. Transfer learning on a pre-trained Convolutional Neural Network (CNN) framework classifies cars based on 196 brands, models, and years utilising minimal data from the Stanford Cars dataset. Fine-tuning yields 85% test accuracy and 96.3 % top-5 test accuracy, surpassing state-of-the-art results. In [16] Donny Avianto et al. classified cars by make and model. Vehicle make and model may assist investigators find traffic violations without
licences plate information. Many methods categorise car types and manufacturers. Similar cars are hard to spot. Due to their similarity, the classifier may make errors. A multi-task learning convolutional neural network fine-grained classifier is suggested. VGG-16 architecture analyses vehicle photographs. Two branches will categorise car model and manufacturer using the data.

3. **CompCars' Dataset Characteristics**

This dataset has been taken from [http://mmlab.ie.cuhk.edu.hk/datasets/comp_cars/index.html](http://mmlab.ie.cuhk.edu.hk/datasets/comp_cars/index.html). All the pictures in this dataset are from public websites and car forums. Surveillance cameras collect images of the surveillance-like type. Real-world applications frequently make use of the data from these two types of study. They allow for cross-modality study of automobiles. This includes 281 car models. Surveillance-nature data includes more than 44,481 front-facing photos of automobiles. It is labelled with a bounding box, model, and colour of the car for each surveillance-nature image. Figure 1 depicts surveillance photographs that have been significantly distorted by lighting and haze fluctuations. As a whole, the CompCars dataset has four unique properties compared to other automobile picture databases: car hierarchy, car attributes, views, and car components.

![Figure 1. Examples of photos from the surveillance-nature data. There are huge differences in the photographs' look due to the changing conditions of light, weather, and traffic, for example.](https://doi.org/10.37965/jait.2023.0268)

From the top to the bottom, as shown in Figure 2, the automobile models can be arranged into a huge tree structure composed of three layers: manufacturer, model, and year of manufacturing. Even more complicated is the fact that automobile models might be built at different times, resulting in minor variations in their appearance. For example, three "Audi A4L" models were made between 2009 and 2011.

**Aspects of a Car:** The maximum speed, displacement, the number of doors, the number of seats, and the type of car are all included in the five qualities assigned to each model. When comparing and contrasting different car models, these characteristics are a gold mine of knowledge. As an illustration, Figure 3 depicts our classification of automobiles, which includes MPVs, SUVs, hatchbacks, sedans, minibuses, fastbacks, estates, pickups, sports cars, crossovers, convertibles, and hardtop convertibles. These qualities can be broken down into two groups: explicit and implicit. Each of these variables is represented by a discrete or continuous value. For example, a door number is represented by a number, whereas a seat number is represented by a letter.
From a correct vantage position, humans can quickly detect how many doors and seats a car has, but they have a hard time identifying the vehicle's top speed and displacement. In addition, for each car type, we designate five distinct points of view: the front (F), the back (R), the side (S), the front (FS), and the rear (FS) (RS). Several expert annotators have dubbed these perspectives. Table 1 displays the distribution of the number of photos of labelled automobiles. Due to difficulties in collecting photographs for some less popular car models, the number of viewpoints for each model is not evenly distributed.

**Parts of a Car:** These photos include four exterior components (headlight, taillight, fog light and air intake) and four interior components (console, steering wheel, dashboard and gear lever) for each model of automobile. For the sake of analysis, the photos have been approximately aligned. Table 2 provides an overview, while Figure 4 illustrates some of the points made.

### 4. APPLICATIONS

Three applications of CompCars are examined here, including fine-grained car classification, attribute prediction and car verification. Select 78 and 126 photos from the CompCars dataset to create three separate subsets without any overlaps. A total of 431 automobiles are included in Part-I, which includes 955 photographs depicting the full vehicle and a further 20349 images depicting individual automobile parts. There are 454 photos in total in the second subset (Part-II) of 111 models. There are 1145 car models with 22236 photos in the final subset (Part-III). Some raw sample images are showed in Figure 5 and processed(cropped) images in Figure 6. Using photos from the first selection, classification of cars at a finer level...
is possible. Models are trained on the first subset and tested on the second for attribute prediction. The final subset is used to verify the car's authenticity.

5. METHODOLOGY
This is where our assessments begin, by comparing the results of many deep neural networks with those of specialist state-of-the-art approaches. After training on the CompCar we are able to correctly identify the vehicle's manufacturer and model using a split of 8144 training photos and 8041 testing images. The model's accuracy was 0.9963 in this scenario, we trained our network, which has 31 residual layers.

We used labelled photographs with the background clipped to train the models in order to further improve our results. In this case, the primary goal is to decrease dataset size for that we have first prepared and create very light feature dataset using binary edge form using canny operation as shown in Figure 7.

For the following applications, we use the Convolutional Neural Network (CNN), which has proven successful in various computer vision tasks, such as object categorization, detection, face alignment, and face verification. The Overfeat [20] model is used specifically for automobile classification and attribute prediction, and it is trained on ImageNet classification tasks [3] before being fine-tuned using the car photos. The fine-tuned model is used as a feature extractor for automobile model verification.

The prepossessed featured dataset has made uniform image size of 256x256 dimension. Our proposed CNN was trained and tested on the same lightweight feature dataset to establish a new baseline for future research in this area.

It is possible to integrate the predictions of many models through lower variance and generalisation error, which can lead to superior predictions, i.e., higher performance than any of the contributing models.
Repeatedly sampling a training dataset and then training a new model is one way to accomplish this.

**A. Augmentation of Data**

We have also used data augmentation techniques to enlarge the pictures in the training data. Before feeding the network its input, this process verifies the quality of the sample picture.

**B. Experimental Setup**

Training and testing on the suggested system was done using personal computers having Windows 10 operating system, 32 GB of RAM, Intel Core i7 processor, and NVIDIA 1050Ti 4GB Graphics Card with MATLAB R2020a with deep learning toolbox. Figure 8 shows the proposed car model prediction using CNN for CompCar Surveillance Dataset. The system has main modules which are as follows

1. **Raw CompCar Dataset**

The dataset we got is has images of different dimensions, but network has not capability to process unequal images. So, this raw dataset needs to be prepared perfectly to get is trained.

2. **Cleaning of Images of Datasets**

All the images are cropped properly to remove unnecessary part of the image except the vehicle and make size of all the images same. Here the image size has been considered is 256x256.

3. **Feature Extraction to reduce size of Dataset**

Network takes lots of time to process colour images, the images present in the dataset has not such colour specific model or make which affects the training process much instead of considering physical profile of the car. So, by reducing the size of input dataset will significantly improve the training time. To reduce the size of the dataset colours of images are processed with binary edge descriptor operation. It gives the physical profile of car or can say that the features of the car.

4. **Training of Network and Validation**

The proposed CNN architecture with 31 layers takes feature sample images as input to get trained. During training process, the network validated itself for validation accuracy (refer to Figure 10), and final accuracy of the network is shown in the graph as well.

5. **Testing**

After completion network training, any test car image can be chosen to get the class (Make/Model) of it.
Figure 8. Proposed Car Model Prediction Framework using CNN for CompCar Surveillance Dataset

The inside layer architecture of the presented CNN is shown in Figure 9. The layer diagram is color coded to understand much better. As you see that the first two layer of the network is 2D conv layers having kernel size of 3x3x16 and 1x1x16 this forms module1 (named ‘conv1’), followed by the maxpool layer followed by two layers of 2D conv layers having kernel size of 3x3x32 and 1x1x32 this forms module2(named ‘conv2’), followed by the maxpool layer. In third module there is one 2d convolutional layer of 3x3x64 kernel and maxpool layer. Followed by two conv layers of size 3x3x64 and 1x1x64, then maxpool layer and one conv layer of size 5x5x64. The sixth module of the network is having 3x3x64 followed by fully connected layer and softmax mentioned with the possible classes(make/models).

6. RESULTS AND DISCUSSION

Figure 10 shows the validation accuracy of the presented CNN network which shows the path of training accuracy to reach to its final stage. This graph is also shows that the presented network has no over fitting and underfitting problem during training and validation phase. Similarly Figure 11 shows the validation loss against iterations and training loss curve.

Figure 9. Proposed CNN Network Architecture

Figure 10. Validation Accuracy of Proposed CNN Network Architecture
This article has been accepted for publication in a future issue of this journal, but it is not yet the definitive version. Content may undergo additional copyediting, typesetting and review before the final publication.

Citation information: Kritika Kashyap, Rohit Miri, An Optimal BDCNN ML Architecture for Car Make Model Prediction, *Journal of Artificial Intelligence and Technology* (2023), DOI: https://doi.org/10.37965/jait.2023.0268

Figure 11. Validation Loss of Proposed CNN Network Architecture

Table 1. Processing Time Performance for testing one Image
<table>
<thead>
<tr>
<th>Model</th>
<th>Testing Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet152</td>
<td>217ms</td>
</tr>
<tr>
<td>InceptionResNetV2</td>
<td>179ms</td>
</tr>
<tr>
<td>Xception</td>
<td>98ms</td>
</tr>
<tr>
<td>DenseNet201</td>
<td>89ms</td>
</tr>
<tr>
<td>DenseNet121</td>
<td>79ms</td>
</tr>
<tr>
<td>MobileNetV2</td>
<td>37ms</td>
</tr>
<tr>
<td><strong>Proposed(this)</strong></td>
<td><strong>40ms</strong></td>
</tr>
</tbody>
</table>
Table 2. Summary of previous work related to vehicle make and model recognition.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Methods</th>
<th>Image Dataset</th>
<th>Outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yong-guo Ren and Shanzhen Lan [11]</td>
<td>develop a CNN-inspired MMR framework for automobiles</td>
<td>Using the CompCars dataset and other sources, we were able to collect 42,624 photos of 233 models.</td>
<td>They achieved an accuracy of 98.7%</td>
</tr>
<tr>
<td>Tafazzoli [12]</td>
<td>VGG and ResNet-50</td>
<td>VMMRdb-3036</td>
<td>Top-level accuracy was 51.76% and top-five accuracy was 92.90%</td>
</tr>
<tr>
<td>Lei Lu and Hua Huang [13]</td>
<td>There are two levels of this hierarchical classifier: brand recognition and model recognition.</td>
<td>12,238 photos of vehicles' front ends were used, representing 400 unique models from 58 different manufacturers.</td>
<td>The final accuracy 95.01%</td>
</tr>
<tr>
<td>Yang et al [14]</td>
<td>Joint Bayesian and CNN</td>
<td>CompCars Dataset</td>
<td>Using full-car photos, they were able to reach an accuracy of 82.2%, 83.0%, and 761 on medium, easy, and hard difficulties, respectively, while the make accuracy was distributed between 59.7% and 82.9%.</td>
</tr>
<tr>
<td>Benavides and Tea [15]</td>
<td>VGG16</td>
<td>Stanford cars dataset</td>
<td>They performed at an 85% level on the test</td>
</tr>
<tr>
<td>Liu and Wang [16]</td>
<td>VGG, CaffeNet, &amp; GoogleNet using transfer learning</td>
<td>Stanford cars dataset</td>
<td>Accuracy of 80% was ultimately reached using GoogleNet</td>
</tr>
<tr>
<td>Our Work</td>
<td>Feature Minimization CNN</td>
<td>CompCars-Surveillance 281 Models and 44,624 images</td>
<td>The optimal Accuracy is 99.25%</td>
</tr>
</tbody>
</table>

Table 2 shows the performance comparison of the previous works with presented work. The accuracy of the previous works is 83% for [11], 92.9% for [12], 85% for [13], 98.7% for [14] and [15] has 95.01% accuracy. Here all the research work carried out with the help of pretrained models like Joint-Bayesian, ResNet-50, VGG, VGG16, GoogleNet, and
CaffeNet etc. But merit of presented work over other networks is having 31 layers optimized CNN. The optimization has been performed by performing frequent tests.

7. CONCLUSION AND FUTURE SCOPE

This research CNN work showed evaluation of the presented CNN against the state-of-the-art network architectures on the same input car dataset and found better accuracy. The key logic to get higher accuracy was extraction of binary descriptor feature using edge detection before training the CNN. This step reduced the size of the car dataset; hence network took less time to get trained. Model outcomes showed that the architecture having 31 layers including 2d convolutional layer, batch normalization, maxpool, ReLU, fully connected layer and Softmax classifier layer, has given 99.25% accuracy. The achieved accuracy is quite near the 100%, which may be ideally not reflected in each and every scenario. But better accuracy showed a step further optimization in the designing process of the network. The feature extraction step prior to input could be replaced with some more efficient segmentation methods to get the key objects extraction out of the input samples. This approach will add some robustness to the prediction reliability to the network and reduces the state of the art pretrained networks complexity and speedup the training and validation process which take sometimes months to train a large network like TensorFlow.

8. REFERENCES


[23] TensorFlow Cars196, Available at: https://www.tensorflow.org/datasets/catalog/cars196