ISTP

An Optimal BDCNN ML Architecture for Car Make Model Prediction

Kritika Kashyap and Rohit Miri

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

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. Analysing pre-existing datasets, we find that the CompCars-SV is 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 analyse 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 normalisation, 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 categorisation, model verification, and attribute prognosis are just a few examples of how the dataset might be put to use.

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

I. 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 is shifting to include more than just functionality and reliability, but also high levels of comfort and style. There are a vast number of automobile designs and models, which makes automobiles a rich object class for computer vision models and algorithms [1]. Cars have a number of unique qualities that other items don't have, which makes categorisation more difficult and allows for a wide range of new research topics. There are a lot of models to choose from in automobiles, allowing for a more demanding fine-grained task than in other categories. Cars' unconstrained postures create considerable appearance disparities, which necessitate viewpoint-aware analysis and algorithms. The car category has its own distinct hierarchy, with three tiers from the top to the bottom: make, model, and year of release. This structure suggests a hierarchical approach to the fine-grained job, which is only described in a small amount of literature [2].

Many fascinating computer vision challenges can be found in automobiles. Firstly, different car manufacturers and different years use distinct design styles, which allows for fine-grained style analysis [3] and fine-grained part recognition. A second reason is because the automobile is a popular subject for attribute forecasting. Furthermore, cars can be identified by their appearance, which includes attributes such as the number of axles and maximum speed and displacement. Car verification, which focuses on determining if two cars are of the same model, is an interesting yet understudied subject in comparison to human face verification [4]. Unrestricted perspectives make car verification more difficult than face verification traditionally.

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 categorisation, 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 [5–7]. 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 is 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 [8–11].

Corresponding author: Kritika Kashyap (e-mail: kritikaphd2022@gmail.com).

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 [12,13]. 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 [14,15]. Additionally, the dataset's multi-scenario structure allows for cross-modality study. Section III 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) [16], to prove its utility and urge the community to investigate new unique research ideas.

The paper is organised into six sections. Section II discusses the related work on vehicle recognition. Section III discusses about the characteristics of the CompCars dataset of cars details and brief about the make and models covered. Section IV discussed about the applications of dataset in developing machine learning models of cars or vehicle recognition. Section V covers the research methodology of this research work to get better accuracy of recognition, the methodology covering the execution step details. Section VI 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.

II. RELATED WORK

In [17], Y. Ren *et al.* discussed about "many smart applications use vehicle-type categorisation (VTC), licence 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 CNNs in this study. First, CNNs were trained and tested on car frontal pictures. Our vehicle MMR identification accuracy is 98.7% using our suggested framework." In [16], F. Tafazzoli *et al.* demonstrate that 'Vehicle Make and Model Recognition (VMMR)' supports 'Intelligent Transportation Systems/Autonomous Vehicle Systems' (ITS/AVS). Real-time and accurate VMMR leads to resource savings. 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 [3], author discussed about "Vehicle-based VMMR. This job demands expertise. Aesthetically, automotive frontal structural components differ; these structures possess distinct traits. Car logos and intra-brand models play a crucial role in shaping brand identity, while class subclasses also share similarities with other courses. Step 2 trains subclass-specific discriminant subregion classifiers. Patches improved this article's big car photo collection". In [18], the study focuses on highlights of 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 datasets, along with web-nature datasets, categorize, verify, and forecast

automotive models. We discuss automobile issues and applications. In [19], Linjie Yang *et al.* show 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 pretrained CNN framework classifies cars based on 196 brands, models, and years utilising minimal data from the Stanford Cars dataset [20]. Fine-tuning yields 85% test accuracy and 96.3 % top-5 test accuracy, surpassing state-of-the-art results. In [2], Donny Avianto *et al.* classified cars by make and model. Vehicle make and model may assist investigators find traffic violations without licence plate information. Many methods categorise car types and manufacturers. Similar cars are hard to spot. Due of their similarity, the classifier may make errors. A multi-task learning CNN fine-grained classifier is suggested. VGG-16 architecture analyses vehicle photographs. Two branches will categorise car model and manufacturer using the data.

III. COMPCARS' DATASET CHARACTERISTICS

This dataset has been taken from *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 include 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 [21,22].

From the top to the bottom, as shown in Fig. 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.
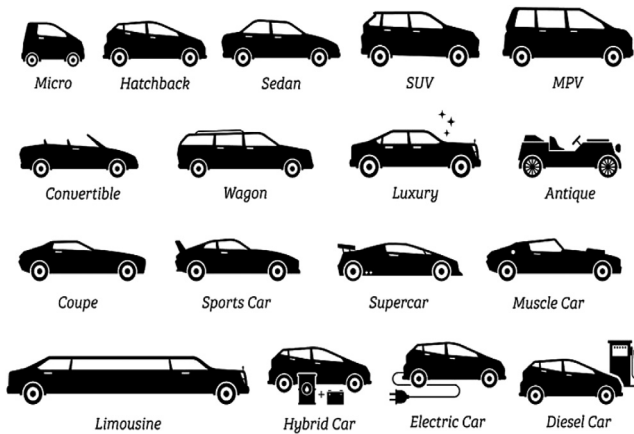
Aspects of a Car: The maximum speed, displacement, the number of doors, the number of seats, and the type of car are all included in the five qualities assigned to each model. When comparing and contrasting different car models, these

Fig. 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.
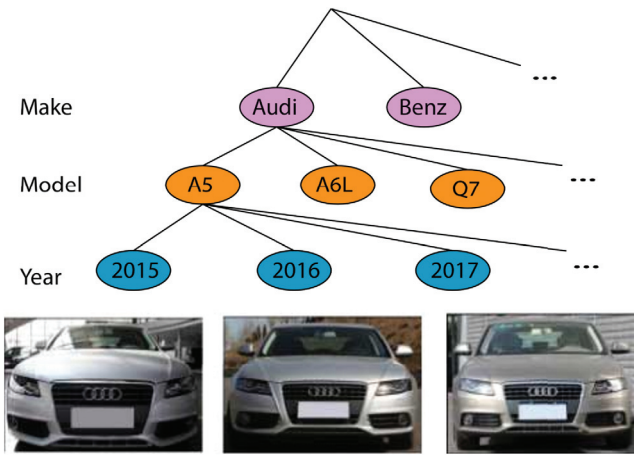
Fig. 2. The tree-like hierarchy of automobile models. There are also a number of Audi A4Ls from various years on show.

Fig. 3. Classification of different car bodies.

Table I. Processing time performance for testing one image

Model	Testing time
ResNet152	217 ms
InceptionResNetV2	179 ms
Xception	98 ms
DenseNet201	89 ms
DenseNet121	79 ms
MobileNetV2	37 ms
Proposed(this)	*40 ms*

characteristics are a gold mine of knowledge. As an illustration, Fig. 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 I 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 II provides an overview, while Fig. 4 illustrates some of the points made.

Table II. Summary of previous work related to vehicle make and model recognition

Ref.	Methods	Image dataset	Outcomes
Yong-guo Ren and Shanzhen Lan [17]	develop a CNN-inspired MMR framework for automobiles	Using the CompCars dataset and other sources, we were able to collect 42,624 photos of 233 models.	They achieved an accuracy of 98.7%
Tafazzoli [16]	VGG and ResNet-50	VMMRdb-3036	Top-level accuracy was 51.76% and top-five accuracy was 92.90%.
Lei Lu and Hua Huang [3]	There are two levels of this hierarchical classifier: brand recognition and model recognition.	12,238 photos of vehicles' front ends were used, representing 400 unique models from 58 different manufacturers.	The final accuracy 95.01%
Yang et al. [18]	Joint Bayesian and CNN	CompCars Dataset	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%.
Benavides and Tea [19]	VGG16	Stanford cars dataset	They performed at an 85% level on the test
Liu and Wang [2]	VGG, CaffeNet, & GoogleNet using transfer learning	Stanford cars dataset	Accuracy of 80% was ultimately reached using GoogleNet
Our Work	Feature Minimisation CNN	CompCars-Surveillance 281 Models and 44,624 images	The optimal Accuracy is 99.25%

Fig. 4. Each column displays 8 car parts from a car model. The corresponding car models are Buick GL8, Peugeot 207 hatchback, Volkswagen Jetta, and Hyundai Elantra from left to right, respectively.

IV. 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

Fig. 5. Raw dataset before pre-processing.

Fig. 6. Formatted (Cropping and Resizing) dataset before pre-processing.

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 Fig. 5 and processed(cropped) images in Fig. 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.

For the following applications, we use the CNN, which has proven successful in various computer vision tasks, such as object categorisation, detection, face alignment, and face verification. The Overfeat [23] model is used specifically for automobile classification and attribute prediction, and it is trained on ImageNet classification tasks [24] before being fine-tuned using the car photos. The fine-tuned model is used as a feature extractor for automobile model verification.

V. 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

Fig. 7. Dataset after pre-processing to optimise the size of dataset.

0.9963 in this scenario, and 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 Fig. 7.

The prepossessed featured dataset has made uniform image size of 256×256 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 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 256×256.

3) FEATURE EXTRACTION TO REDUCE SIZE OF DATASET. Network takes lots of time to process colour images, and 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, reducing the size of input dataset will significantly improve the training time. To reduce the size of the dataset, the colors of the images are processed using a binary edge descriptor operation. This operation provides the physical profile of the car, or in other words, 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 Fig. 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.

The inside layer architecture of the presented CNN is shown in Fig. 9. The layer diagram is colour-coded to understand much better. As you see, the first two layers of the network are 2D convolutional layers with kernel sizes of $3 \times 3 \times 16$ and $1 \times 1 \times 16$. This forms module1 (named 'conv1'), followed by the maxpool layer. This is followed by two layers of 2D convolutional layers with kernel sizes of $3 \times 3 \times 32$ and $1 \times 1 \times 32$. This forms module2 (named 'conv2'), followed by the maxpool layer. In the third module, there is one 2D convolutional layer with a $3 \times 3 \times 64$ kernel and a maxpool layer. This is followed by two convolutional layers with sizes of $3 \times 3 \times 64$ and $1 \times 1 \times 64$, then a maxpool layer, and one convolutional layer with a size of $5 \times 5 \times 64$. The sixth module of the network has a $3 \times 3 \times 64$ configuration followed by a fully connected layer and Softmax, which is mentioned along with the possible classes (make/models).
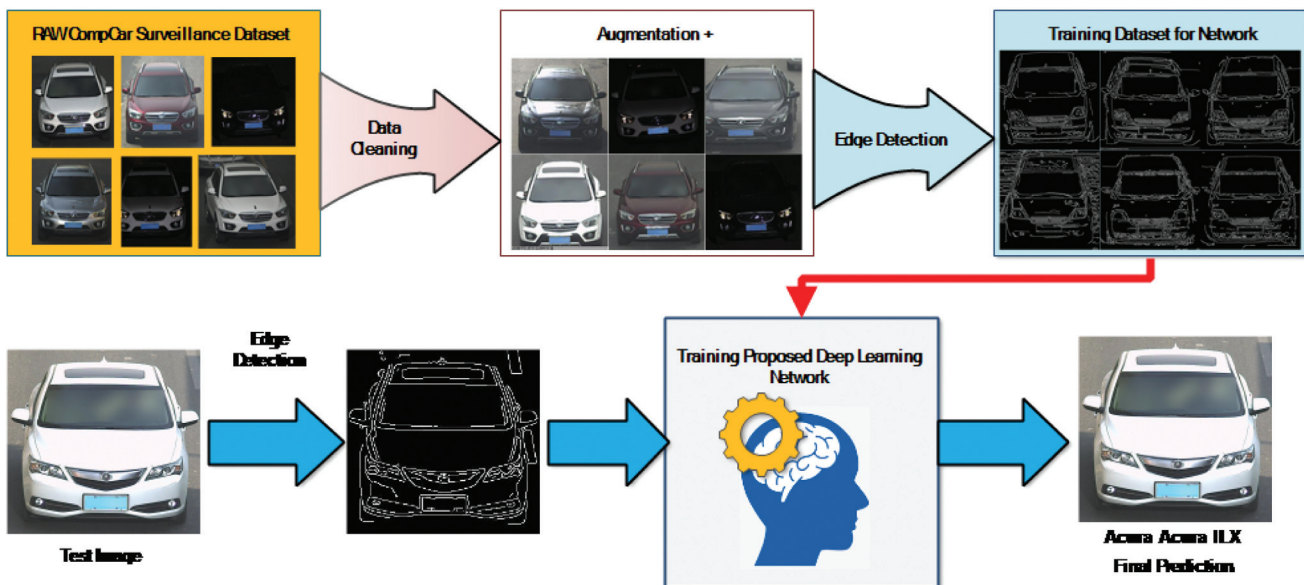

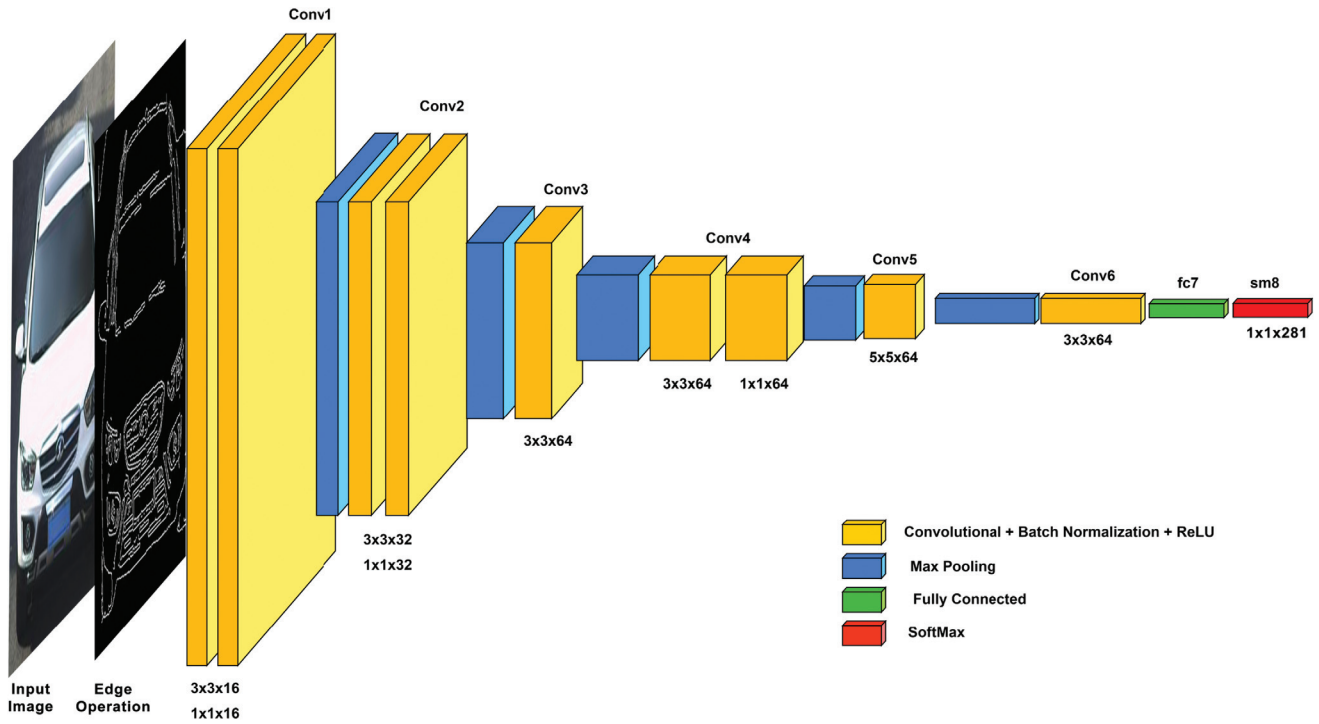**Fig. 8.** Proposed car model prediction framework using CNN for CompCar surveillance dataset.

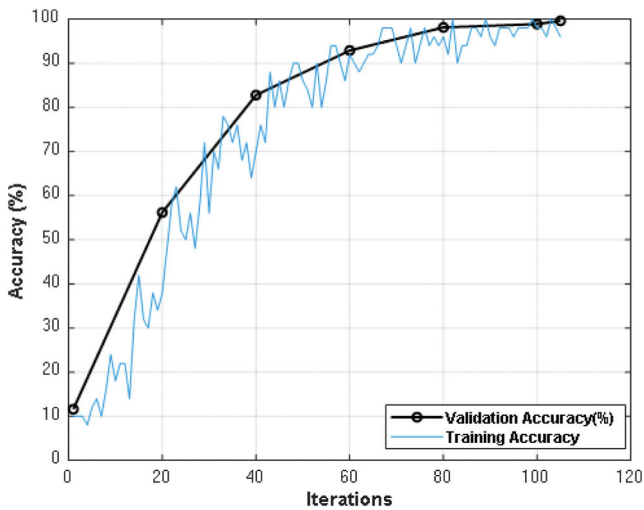**Fig. 9.** Proposed CNN network architecture.

Fig. 10. Validation accuracy of proposed CNN network architecture.
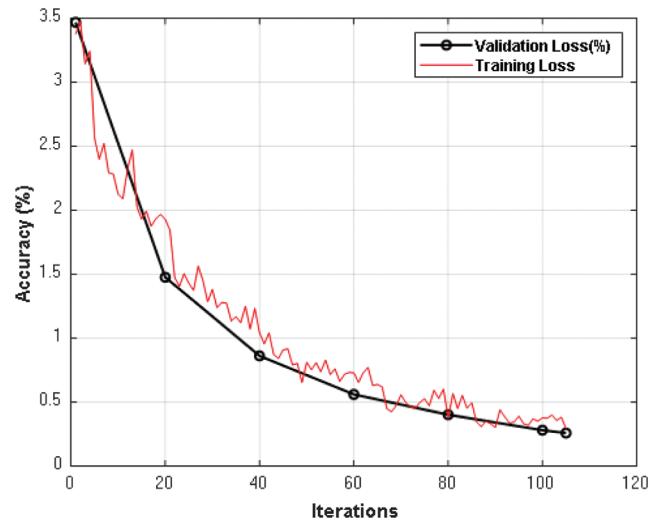

Fig. 11. Validation loss of proposed CNN network architecture.

VI. 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 also shows that the presented network has no over fitting and underfitting problem during training and validation phase. Similarly, Fig. 11 shows the validation loss against iterations and training loss curve.

Table II shows the performance comparison of the previous works with presented work. The accuracy of the previous works is 83% for [17], 92.9% for [16], 85% for [3], 98.7% for [18,19] and

has 95.01% accuracy. Here, all the research works were 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 optimised CNN. The optimisation has been performed by performing frequent tests.

VII. CONCLUSION AND FUTURE SCOPE

This research 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 normalisation, 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 optimisation 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 reduce 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.

CONFLICT OF INTEREST STATEMENT

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

REFERENCES

[1] J. MacCarthy, "Vehicle surveillance and control system," *US7548803B2*, Jun. 16, 2009 Accessed: Nov. 04, 2022. [Online]. Available: https://patents.google.com/patent/US7548803/en

[2] D. Liu and Y. Wang, "Monza: image classification of vehicle make and model using convolutional neural networks and transfer learning," Stanford University: Stanford, CA, USA, 2017. [online]. Available: http://vision.stanford.edu/teaching/cs231n/reports/2015/pdfs/lediurfinal.pdf

[3] L. Lu and H. Huang, "A hierarchical scheme for vehicle make and model recognition from frontal images of vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1774–1786, May 2019. DOI: 10.1109/TITS.2018.2835471.

[4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: a large-scale hierarchical image database," in *2009 IEEE Conf. Comput. Vis. Pattern Recogn.*, Jun. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[5] H. Huang and S. Lin, "WiDet: Wi-Fi based device-free passive person detection with deep convolutional neural networks," *Comput. Commun.*, vol. 150, pp. 357–366, Jan. 2020. DOI: 10.1016/j.comcom.2019.09.016.

[6] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016. DOI: 10.48550/ARXIV.1605.07146.

[7] T. Begin, A. Busson, I. Guérin Lassous, and A. Boukerche, "Video on demand in IEEE 802.11p-based vehicular networks: analysis and dimensioning," in *Proc. 21st ACM Int. Conf. Model. Anal. Simul. Wireless Mobile Syst.*, New York, NY, USA, Oct. 2018, pp. 303–310. DOI: 10.1145/3242102.3242109.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. DOI: 10.48550/ARXIV.1409.1556.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *2016 IEEE Conf. Comput. Vis. Pattern Recogn. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778. DOI: 10.1109/CVPR.2016.90.

[10] J. Krause, H. Jin, J. Yang, and L. Fei-Fei, "Fine-grained recognition without part annotations," in *2015 IEEE Conf. Comput. Vis. Pattern Recogn. (CVPR)*, Boston, MA, USA, Jun. 2015, pp. 5546–5555. DOI: 10.1109/CVPR.2015.7299194.

[11] B. Zhao, X. Wu, J. Feng, Q. Peng, and S. Yan, "Diversified visual attention networks for fine-grained object classification," *IEEE Trans. Multimed.*, vol. 19, no. 6, pp. 1245–1256, Jun. 2017. DOI: 10.1109/TMM.2017.2648498.

[12] 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. 91487–91499, 2021. DOI: 10.1109/ACCESS.2021.3090766.

[13] D. Long, R. Zhang, and Y. Mao, "Prototypical recurrent unit," *Neurocomputing*, vol. 311, pp. 146–154, Oct. 2018. DOI: 10.1016/j.neucom.2018.05.048.

[14] K. Cho *et al.*, "Learning phrase representations using RNN encoder-decoder for statistical machine translation," 2014. DOI: 10.48550/ARXIV.1406.1078.

[15] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3D object representations for fine-grained categorization," in *2013 IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2013, pp. 554–561. DOI: 10.1109/ICCVW.2013.77.

[16] F. Tafazzoli, H. Frigui, and K. Nishiyama, "A large and diverse dataset for improved vehicle make and model recognition," in *2017 IEEE Conf. Comput. Vis. Pattern Recogn. Workshops (CVPRW)*, Jul. 2017, pp. 874–881. DOI: 10.1109/CVPRW.2017.121.

[17] Y. Ren and S. Lan, "Vehicle make and model recognition based on convolutional neural networks," in *2016 7th IEEE Int. Conf. Softw. Eng. Serv. Sci. (ICSESS)*, Aug. 2016, pp. 692–695. DOI: 10.1109/ICSESS.2016.7883162.

[18] L. Yang, P. Luo, C. C. Loy, and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," 2015. DOI: 10.48550/ARXIV.1506.08959.

[19] C. T. N. Benavides, "Fine grained image classification for vehicle make and model using convolutional neural network," Stanford Univ., Stanford, CA, USA, vol. Tech. Rep. 18681590, 2019, [Online]. Available: http://cs230.stanford.edu/projects_spring_2019/reports/18681590.pdf

[20] TensorFlow Cars196. Available: https://www.tensorflow.org/datasets/catalog/cars196.

[21] X. Liu, T. Xia, J. Wang, Y. Yang, F. Zhou, and Y. Lin, "Fully convolutional attention networks for fine-grained recognition," 2016. DOI: 10.48550/ARXIV.1603.06765.

[22] J. Krause, M. Stark, J. Deng, and L. Fei-Fei, "3D object representations for fine-grained categorization," *4th IEEE Workshop 3D Represent. Recogn*, at ICCV 2013 (3dRR-13). Sydney, Australia. Dec. 8, 2013. Available: https://ai.stanford.edu/~jkrause/cars/car_dataset.html

[23] A. Paszke *et al.*, "Automatic differentiation in PyTorch," Oct. 2017. [Online]. Available: https://openreview.net/forum?id=BJJsrmfCZ. Accessed on Nov. 04, 2022.

[24] Q. Hu, H. Wang, T. Li, and C. Shen, "Deep CNNs with spatially weighted pooling for fine-grained car recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 3147–3156, Nov. 2017. DOI: 10.1109/TITS.2017.2679114.