2021 6th International Electrical Engineering Conference (IEEC 2021)
April, 2021 at NEDUET, Karachi, Pakistan

Fine-Grained Classification of Vehicles by using Convolutional Neural Network (CNN)

Danish ul Khairi[1], Mohammad Umair Arif[2], Mohammmd Sharjeel Habib[3] and Syed Mamnoon Akhter[4]
[1,3] Department of Electrical Engineering, Bahria University
Email: [1]danishulkhairi@gmail.com, [2]Umairarif.bukc@bahria.edu.pk,, [3]Sharjeelhabib98@gmail.com,
[4]Applied Physics, University of Karachi
Email: [4]Mamnoon.eng@gmail.com

Abstract: Machine Learning has a practical and profound application in intelligent traffic management systems. ITS is a very broad terminology in which includes vehicle detection, classification, monitoring, surveillance, license plate recognition, etc. Vehicle classification playing a vital role in the intelligent transportation system for traffic management and traffic monitoring. The aim of this study is the fine-grained classification of vehicles using convolutional neural networks. To accomplish the task there are lots of challenges involved. Some of them are Inter-class and Intra-class similarities between the make and models of vehicles, lightning conditions, background, shape, pose, viewing angle of the camera, vehicle size, color occlusion, and environmental conditions. There are three different datasets used in this paper BMW-10, Stanford Cars, and PAKCars.The BMW-10 and Stanford Cars datasets are available open-source, where the PAKCars dataset is self-generated mainly for the fine-grained classification of cars in Pakistan to analyze the implementation of research. In the training part of the system, three different DCNN models are Inception-V3, VGG-19, and ResNet-50 used. Each model trained on all three datasets (BMW-10, Stanford Cars, and PAKCars). A total of 10 classes are evaluated in the BMW-10 dataset having a total of 511 images while 196 classes are evaluated in Stanford Cars datasets having 8144 training images and 44 classes evaluated in PAKCars datasets which have a total of 1000 images. The results acquired after processing reveals the performance of true classification ResNet-50, VGG-19, and inception-V3 respectively. VGG-19 and Resnet-50 are more accurate, because of their higher numbers of layers and architecture that make them complex and more computational power consuming as compared to Inception-V3.

Keywords: Inception-V3, VGG-19, Resnet-50, Track surveillance, Deep Neural Networks, PakCars, Intelligent Transportation System (ITS)

I. INTRODUCTION

Artificial intelligence, machine learning, and computer vision have become popular among researchers, students, and professionals currently. Computer vision can be defined as helping or enabling a computer to see just like a human does it? Computer vision is an advanced form of image processing. It involves the areas of computer science, mathematics, and electrical engineering. Computer vision includes image acquisition, the process to understand the images or videos, and the extraction of useful features and information from images or videos to mimic the human vision. A human can easily see grayscale images and RGB images but the computer vision can gather additional information from images like height, pixels information, edges, segmentation, etc. Computer vision can also be used to analyze and process depth and IR images.

Computer vision and artificial intelligence share many topics like image processing, pattern recognition, and machine learning. To mimic the human vision researchers develop lots of learning algorithms to embed human vision into a machine. Researchers are finding that how humans can identify objects? How the vision is strong? Which features the human use to get such strong performance on vision tasks? Normally computer vision algorithms extract the features from images in the form of image vectors and use them to make decisions.

Object detection, object classification, object recognition, and speech recognition are some important tasks performed using computer vision. Traditional popular computer vision algorithms are SIFT (scale-invariant features transformation)[1] and SURF (speeded-up robust features) which extract the features from images[2]. Now Deep Convolutional Neural Networks being more accurate are replacing the above methods and are widely used to perform such tasks[3].

II. LITERATURE REVIEW

Deep learning has become a popular area among researchers now. Recent developments in deep learning have significantly improved the performance of computer vision in many application areas [4]. In[5] the author presents the application of Channel Max Pooling Modified CNNs to perform Fine-Grained Vehicle classification. The generalization ability of CNN is improved by learning more discriminative features using fewer feature maps. In this research, a new layer was introduced between fully connected layers and convolution layers. Channel max-pooling eliminates redundant information by conducting the Max Pooling operation along the channel side among the corresponding positions of the consecutive feature maps. Results obtained by applying this scheme on two prominent datasets (Stanford Cars-196 and CompCars) validated the effectiveness of the method with a relatively less number of parameters. Results show that the performance of DenseNet-161, VGG-16, and ResNet152-WoGAP was improved using channel max pooling.

In this paper, [6] author discusses the application of fine-grained vehicle recognition in the field of traffic monitoring and related applications. In this research work approach of 3D bounding boxes was proposed that enables the user to normalize the image by unpacking in the form of a plane. This method was found very efficient in improving accuracy as well as reducing errors. It improved accuracy by 12% while classification errors were reduced by 50% relative to the base CNN architecture used in this research. A large-scale dataset named BoxCars116k was also collected that contains 116K images of vehicles taken from different viewpoints. The use of a 3D bounding box technique eliminates the need for images captured from the front or rear viewpoints or the use of a 3D model that makes this method suitable for large-scale deployment. A new dataset BoxCars116k was collected and annotated by authors, containing 116,286 images of 27,496 vehicles which are of 45 different makes with 693 classes, captured using 137 different cameras and different viewpoints. The following are the nets with proposed modifications that were used AlexNet, VGG(16 and 19 layers), ResNet(50, 101, and 152 layers) DCNN with a compact bi-linear pooling layer.

In this paper, [7] latent support vector machines were used to perform fine-grained vehicle recognition. Also, a part-based model is learned for each category. Authors have proposed a novel training technique in which different parts of cars are cropped from images and include in a dataset for part-based classification of vehicles. To make the system faster a unique cascading scheme was introduced that classifies the input images sequentially according to their Frequency and Confidence. Vehicle recognition has many additional challenges like visual features and distinct structural, a small inter-class distance, and a large intra-class variation. To overcome these challenges this research proposed a new cascading scheme, a customized model training procedure, and a parts localization algorithm. A combination of standard SVM and LSVM was used to train a model per class as well as in finding the discriminative parts of each class. A new cascading scheme was proposed to reduce the system processing. This scheme is capable of controlling the tradeoff between accuracy and speed by using the classifiers sequentially. It was concluded that the proposed system is 80% faster with a little bit reduction in overall accuracy. This research work obtained 97.01% on the BVMMR dataset while that of 95.55% achieved on the CompCars dataset.

Vehicle recognition and classification is a challenging dilemma for an intelligent and smart transportation system due to vague and indefinite ‘intra category’ variations in the appearance of car models of the same makes. In [5] this paper author proposed the mentioned problems that can be tackled by locating and identifying discriminative parts where the most obvious variation appears. Using Convolutional Neural Network the discriminative regions are identified automatically. The results surpass most of the other techniques and reached an accuracy of 98.2% over 281 vehicle makes and models.

In [8] Fine-grained vehicle recognition system that is meant to retrieve information related to the vehicles, embedded in the system with the help of training part of machine learning. The particulars retrieved will be the category information of a vehicle/car including; car make, model, and its year of manufacturing. The basic purpose of Fine-grained car classification to recognize a vehicle among the categories and subcategories within the same vehicle model. The results of the above paper showed an improvement in the accuracy from 91.2% to 97.6% on the latest dataset of CompCars and from 92.6% to 93.1% on the dataset of Stanford Cars-196. According to research, the result showed that the SWP (Spatially Weighted Pooling) technique improved the precision of fine-grained classification of cars. For example; VGG-16 and Res-Net101 achieved 85.4% and 90.9% accuracy for the Cars-196 dataset while after using SWP the performance was improved to 90.7% and 93.1% respectively. Also, Res-Net 101 along with the SWP layer excelled in other modern approaches and therefore reported the best results for dataset Cars-196 and CompCars.There are many approaches to car model classification. In this research car model classification can be divided into three main categories namely: texture-feature-based approach, 3D representation-based approach, and DCNN-based approach.

[9] This paper focuses on the Principal Component Analysis Network-based Convolutional Neural Network vehicle recognition system. This technique used only one local feature of a vehicle for recognition of the model. That local discriminative feature was headlamp used to determine the model. The research suggested a model that purges the need for locating and segmenting the headlamps accurately. The model PCNN establishes the usefulness of both the PCA and CNN in extorting the features from a vehicle’s headlamp image and reduces the complexity of CNN. The model is capable of addressing the issue of different types of distortions in an image. The outcomes of the research are dramatic in achieving the accuracy rate of 99.51% on 38 vehicle makes and models. CompCar dataset was used to validate the effectiveness of the model with an accuracy rate of 89.83% on 357 vehicle models.

[9] This paper focuses on the Principal Component Analysis Network-based Convolutional Neural Network vehicle recognition system. This technique used only one local feature of a vehicle for recognition of the model. That local discriminative feature was headlamp used to determine the model. The research suggested a model that purges the need for locating and segmenting the headlamps accurately. The model PCNN establishes the usefulness of both the PCA and CNN in extorting the features from a vehicle’s headlamp image and reduces the complexity of CNN. The model is capable of addressing the issue of different types of distortions in an image. The outcomes of the research are dramatic in achieving

the accuracy rate of 99.51% on 38 vehicle makes and models. CompCar dataset waIn this paper [10] author focus was to get the unsupervised feature learning methods incorporated, for which it takes the input SIFT (scale Invariant Feature Transform) and encoded by LLC (Locality constraint Linear Coding) method for fast encoding. The model used in the study can identify and classify 50 models of vehicles. The model is efficient enough to detect vehicles other than those 50 models as 'unknown vehicle(s)'. The study used two datasets for examining the framework, the first one is the CompCars and the second one is the Iranian on-road vehicle dataset. The results of the study concluded comparable outputs as 98.4% and 97.5% of accuracies.

This [11] paper uses CNN which used two steps: pre-training and fine-tuning respectively. For pre-training, GoogLeNet is pre-trained on dataset ILSVRC 2012. At the second stage of fine-tuning, the model fine-tunes the vehicle dataset for classification. The vehicles were divided into six different categories such as a car, bus, motorcycle, minibus, van, and truck. The results showed that the classification percentage was around 98.26% which is comparatively 3.42% higher than the traditional and conventional method with the use of Feature + Classifier.

The research [12] has suggested a model that is based on a deep convolutional neural network system for recognizing the car make, model, and color. The system is found to be economical and gives better results. The MMCR (Make-Model-Colour-Recognition) is of great opportunity for new domains concerning traffic control systems such as the enforcement of the law, surveillance, driver assistance, and traffic monitoring. The study method was trained to detect 59 vehicles of different make and 818 models. At first, the data is collected then processing is done on collected data. Once data is processed the system is trained. In the data collection stage, images are collected from different sources. The color was then recognized in the study. A set of 10 colors was labeled such as; black, white, blue, red, orange, purple, green, grey, yellow, and beige. In the second stage, data is pre-processed. The alignment of images means the labeled vehicles are centered in the image through the 'Sight Hound' model detector. The alignment of images reduces the effect of the background using the vehicle bounding boxing method. In this feature to avoid inaccuracy, a 10% extra margin is selected around the vehicle box. In the last stage of training i.e. the deep training includes two deep neural networks out of which one goes for make and model recognition and the second one is for color recognition and classification. The images are labeled at 150 fps.

The research [13]focused on the residual connections on combining the architectures of Inception networks. The new reorganized architecture of residual and non-residual is also worked in inception networks. In a group of Inception-v4 and three residual networks, the result was 3.08% on the test set of ImageNet. The combo of residual connections and Inception architecture concluded that the residual connections are fundamentally necessary for the deep training as Inception networks themselves are very deep. Also, the Inception network without residual connection is made efficient by making it wider and deeper for which a new network Inception-v4 was used as it has more modules than Inception-v3 and is more simplified as well. The study used momentum with a 0.9 decay and gradient clipping for stabilizing the training.

In [14] this paper study aims at VTR (Vehicle Type Recognition) with improving the rate of accuracy when it comes to filtering images from multiple viewpoints. For the said problem Feedback enhancement Multi-branch Convolutional Neural Network (FMCNN) model is used. The author designed the new dataset MVVTR (Multi-View Vehicle Type Recognition) for testing the effectiveness of FMCNN. CompCars dataset was used for fine-grained vehicle recognition and the results showed 94.9% accuracy in coarse-grained vehicle recognition for MVVTR and 97.8% and 91.0% on CompCars for fine-grained vehicle recognition. FMCNN does not need info of images like viewpoint or camera parameters nor complex pre-processing like part detection or scene segmentation or splitting is used to validate the effectiveness of the model with an accuracy rate of 89.83% on 357 vehicle models.

III. SYSTEM SETUP

In this study, famous convolutional neural network models (Inception, Resnet, and VGG) are used for fine-grained classification of vehicles. These models are ImageNet challenge winners. Deep convolutional neural networks are complex in computational needed to use the high configuration GPU. For this research purpose ACCER Predator Helios 300 laptop having 16GB ram, Core i7 processor, and 6GB Nvidia graphical processing unit GTX 1060 is used. In this work following toolkit are used which are listed below:

- Ubuntu 16.04 (Satble Version)
- Pyhton 2.7 and 3
- CUDA 8.0.4 use for GPU
- Tensorflow version 1.7
- Mxnet Deep convolutional library for convolutional neural Network
- Gluon library for convolutional neural Network
- CUDN used for CUDA used to run faster convolutional Neural Network.

IV. METHODOLOGY

This research aims to implement the fine-grained vehicle classification using different CNN architectures on different datasets. The process of methodology can be divided into two steps training process using CNN architecture and testing process. The datasets we used for our system are Stanford Cars, BMW-10, and PAKCars. A new dataset PAKCars is made and used in research work. It enables us to utilize this research for Pakistan effectively. We made some necessary changes in an available dataset (Stanford Cars and BMW-10 dataset) like cropping the raw image from the given bounding box, sorting images with their respective make and model folders using given labels for both training and testing images. Also, a new dataset PakCars was made specially to be used in research work where we divided each car brand and their respective car models into different folders.

After this, some pre-processing is performed on datasets before training the CNN models. In pre-processing image augmentation is done which includes flipping, Rotation, Zooming, and conversion of an image into black and white. Now the datasets are ready to train on different CNN models (InceptionV3, VGG16, and ResNet). We trained different models using preprocessed data and obtained different outcomes and validate the results. Workflow of methodology is shown in figure 1 and testing process in Fig 2

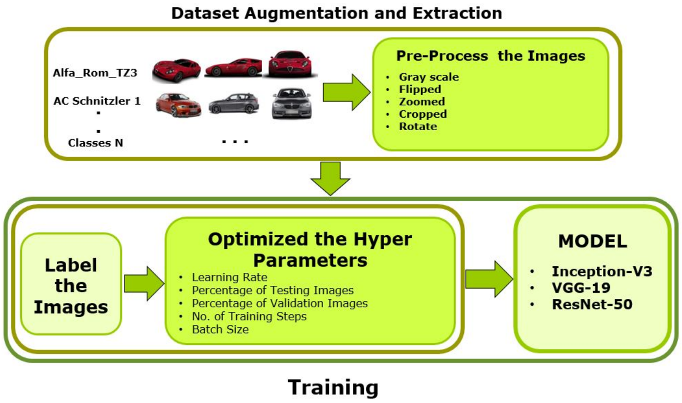

Fig. 1. Methodology of Training process

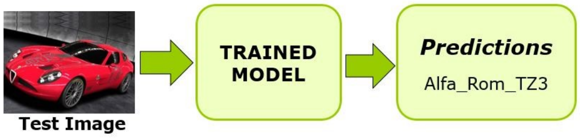

Fig. 2. Methodology of Testing Process

A. Inception-V3

The inception-V3 architecture was first proposed in [15] this paper, which is a widely used model for image classification or recognition. The novelty of the inception architecture is the Inception module. Inception module is consist of multiple parallel convolution filter layer which are 1x1 and 3x3. The main reason to use different filters is to extract the local features with small kernels and feature with context through a larger kernel.

In the Fig 3 shows unveils the conical inception modulo. It also reduces the number of convolution to a maximum of 3x3. Notice that on the left side of an image two 3x3 kernels are stacked together which reduces the number of parameters as compared to a single 5x5 kernel. The weight ratio is $((3 * 3 + 3 * 3)/5 * 5) = 18/25$. After the input 1x1 filter is used to reduce the complexity in computation after the input. The data feed through each branch and then it concatenated as one. This technique made the model fast and accurate, the complete architecture of Inception-V3 as shown in Fig 4.

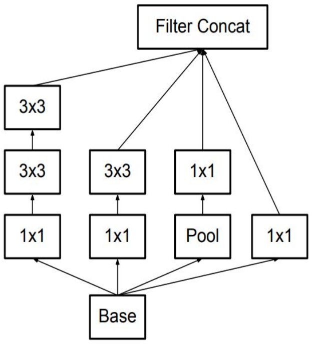

Fig. 3. Inception module [16]

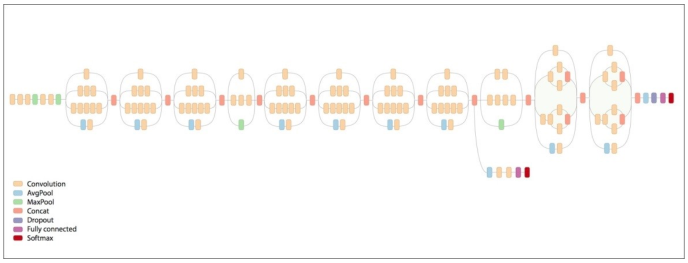

Fig. 4. Complete Architecture of Inception-V3 [17]

B. VGG-19

The VGG architecture is the modified form of AlexNet, it was developed in 2014 by a virtual geometry group. VGG network uses the sequence of 3x3 convolution filter instead of 5x5 and 7x7 single convolution filter. Which gives the same effective receptive field as compared to the single large convolution layer(5x5 or 7x7) and also uses fewer parameters. Many other models were investigated but the VGG performing well as compared to their predecessors. There are two major flaws of VGG 1) Fully connected layer has a hundred million parameters, which increases the computation complexity. This causes a slow training process due to a large number of calculations.2) It uses the same filter size throughout the network design which causes some features to be overlooked. On the other hand, due to the large numbers of parameters model was successfully trained on small datasets for fine-grained classification. The complete architecture of VGG-19 is shown in Fig 5.

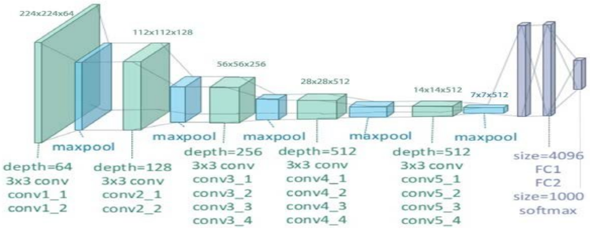

Fig. 5. Layer by layer architecture of VGG-19 [18]

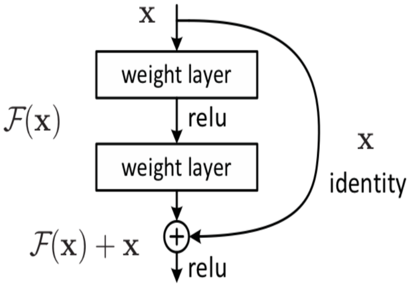
x
weight layer
relu
weight layer
$\mathcal{F}(\mathbf{x})$
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Fig. 6. Residual block of ResNet [19]

C. ResNet-50

ResNet-50 is a classical convolution neural network with 50 layers deep network. It is introduced by Microsoft. It is a smaller version of ResNet 152 and mainly used as a starting point for transfer learning. It is trained on more than a million images from the ImageNet database. It is used as a backbone for many computer vision tasks. This network is 50 layers deep. It can classify images into 1000 object categories, such as a keyboard, pencil, and many animals. It has learned good feature representations for a wide collection of images. It has an image input size of 224-by-224. This network introduces residual learning. In a deep convolution neural network, several layers are stacked and are trained, which learns several level features at the end of its layers. In residual learning, instead of features, we try to learn through residual. Or in other words, Residual is used as subtraction of features learned from the input of that layer. ResNet use shortcut connections that connecting input of the nth layer to some (n+x)th layer directly. It is easier than deep convolutional neural networks. This also resolved the degrading accuracy problem. The ResNet-50 model consists of 5 stages. Each stage has a convolution and Identity block. Each convolution and identity block also has 3x3 convolution layers. The ResNet-50 has over 23 million trainable parameters. The logical scheme of the base building block for ResNet is shown in Fig 6.

V. DATASET PREPRATION

Dataset augmentation is a technique to enhance the training dataset by creating modified versions of images. Training of convolution neural network models with dataset improves the model learning ability. Basic augmentation techniques are super simple and could easily be implemented on any type of image dataset. Some most common basic techniques used for data augmentation are Flip, Rotate, Zooming, Scaling, Cropping, Fill, and Adding Noise. Dataset augmentation is shown in Fig 7. The datasets are used in this research are Stanford Cars, BMW-10 Cars, and PAKCars dataset. For extraction of images first, we used the Stanford Cars dataset. The dataset has been generated by web images these images are collected from Google, Amazon, Bing, etc. where noise-free pictures are included in the dataset for proper training of models on different convolution networks. The dataset contains 196 car models of different makes. In general classes may categories as SUVs, Coupes, Hatchbacks, Pickups, Convertible, Sedan, and Station Wagon. Stanford Cars dataset includes test and train annotation files, cars meta (cars makes) file Training, and testing images folders. In the annotation file, the bounding box of all the cars is available according to the image's name. We use the bounding box points from the given file to crop all images and sort all images into classes. The processed images are shown in Fig 8. Another dataset BMW cars is used in this research work, same process has been implemented to extract the images from the dataset because the BMW-10 cars dataset is the part of Stanford Cars dataset it includes the annotation files same as the Stanford cars dataset which provides the bounding box information of images and class ID of images. BMW-10 cars dataset only consists of different models of BMW. Th PAKCars dataset is generally made for Pakistan, for fine-grained vehicle classification, it is a

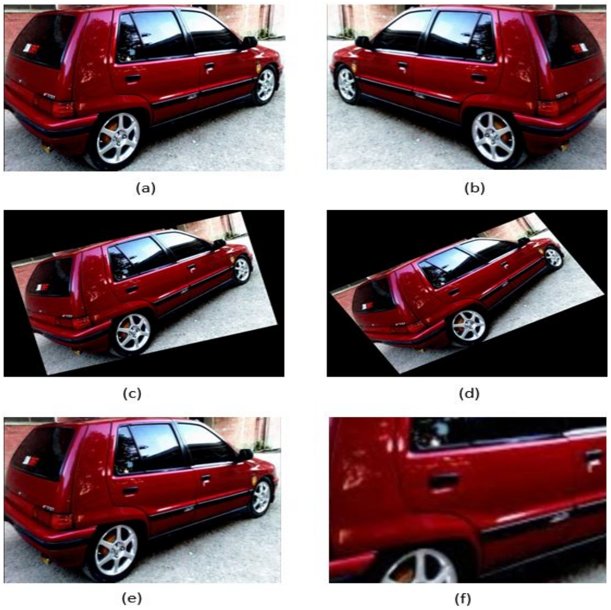

Fig. 7. (a) Original Image (b) Flip Image (c) Rotated 30 Degree (d) Rotated 30 Degree (e) Resized Image and (f) Zoomed Image

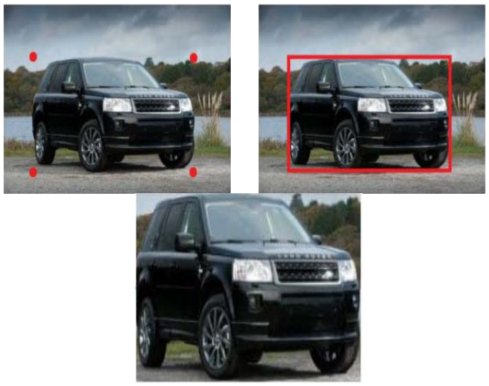

Fig. 8. fig:subfig1 Image with bounding box points fig:subfig2 Bounding box created on image and fig:subfig3 Cropped image

TABLE I
OBTAINED RESULT ACCURACY CHART WITH RESPECT TO DATASETS AND MODELS

Inception-V3							
Datasets	**Training Images**	**Testing Images**	**No. of Classes**	**Learning Rate**	**Training Steps or Epochs**	**Model Parameters**	**Final Test Accuracy**
PakCars	1186	1175	44	0.01	20000	23M	66
BMW-10	245	244	10	0.01	20000	23M	74.6%
Stanford Cars	8144	8041	196	0.01	20000	23M	62%
VGG19 Model							
Datasets	**Training Images**	**Testing Images**	**No. of Classes**	**Learning Rate**	**Training Steps or Epochs**	**Model Parameters**	**Final Test Accuracy**
PakCars	1186	1175	44	0.01	100	143M	84%
BMW-10	245	244	10	0.01	100	143M	87%
Stanford Cars	8144	8041	196	0.01	100	143M	78%
ResNet 50 Model							
Datasets	**Training Images**	**Testing Images**	**No. of Classes**	**Learning Rate**	**Training Steps or Epochs**	**Model Parameters**	**Final Test Accuracy**
PakCars	1186	1175	44	0.01	100	25.6M	83%
BMW-10	245	244	10	0.01	100	25.6M	85%
Stanford Cars	8144	8041	196	0.01	100	25.6M	82%

small-scale dataset. Images are included from different angles and viewpoints. It helps in the detection and fine-grained classification of vehicles in Pakistan. The dataset is taken from different websites like Olx, Pakwheels, and Google images, etc. The dataset contains 44 classes of cars that are commonly used on Pakistani roads.

TABLE II
COMPARISON OF OBTAINED RESULTS WITH PREVIOUS WORK

Stanford Cars Dataset		
Research Papers	**CNN Models**	**Accuracy %**
2*[20]	VGG19	62.5
	Inception-V3	67.5
2*[21]	VGG-19	79.20
	ResNet-50	78.87
[22]	VGG-19	78.90
3*Ours	Inception-V3	62
	VGG_19	78
	ResNet-50	82
BMW-10 Cars Dataset		
[23]	VGG-19	78.74
Ours	VGG-19	87

VI. RESULT AND DISCUSSION

The experimental results of DCNN models on each dataset are shown the table I. The comparison of obtained results with previous work is presented in another table II. In this research, three different datasets (Stanford Cars, BMW-10, and PAKCars) are used for the fine-grained classification of vehicles. Many researchers have trained multiple DCNN models on the Stanford Cars dataset and BMW-10 Dataset. The results obtained in this research work are compared with those obtained by earlier researchers in the table II. In [22] paper researcher used 50,000 to 250,000 epochs with fine-tunes. In this paper[20] researchers use the Keras package for the training of models, and to prepare the dataset they used Keras image data generator to enhance the dataset. Although the accuracy obtained is slightly inferior, it is worth mentioning that our proposed algorithm for data augmentation produces much fewer images, which reduces the training time. Due to the limitations of hardware setup, it is unfortunate that models in this study are limited to be trained under 100 epochs (VGG and ResNet) and 20000 epochs for Inception-V3. However, in the future, it is possible to conduct training with more epochs so that we can understand the potency of each model better and improved performance can be obtained.

VII. CONCLUSION

Vehicle classification playing a vital role in the advancement of ITS systems. This task is based on machine learning and computer vision which helps in developing Urban and SMART cities by examining the traffic and assist in preventing. Fine-grained classification of vehicles helps in many ways in ITS systems. Although, the outcome of this study is satisfactory there is a lot of potentials to improve and extend the algorithm presented in this study. It was concluded that variation in lightening conditions affects the accuracy of the system so one direction could be further improving our classifier by training the classifier on various illumination conditions. The addition of more models of cars will increase the number of classes in PakCars, which will be very helpful for the improvement of fine-grained classification. The results of these models can further be improved by using ensemble CNN by combining the statistical results of multiple models for a single prediction. Furthermore, other models can be added to the comparison list of models to make the comparison more reliable and vast. Part-based classification can lead to eliminating the false classification of damaged cars. The proposed study classifies stored images of vehicles that are not of much significant use. There is a challenging task to implement this technique in real-time so it can be utilized in surveillance and navigation more efficiently.

REFERENCES

[1] M. A. Manzoor and Y. Morgan, “Vehicle make and model classification system using bag of sift features,” in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*, pp. 1–5, Jan 2017.

[2] S. Anuja Prasad and L. Mary, “A comparative study of different features for vehicle classification,” in *2019 International Conference on Computational Intelligence in Data Science (ICCIDS)*, pp. 1–5, Feb 2019.

[3] H. Venkateswara, S. Chakraborty, and S. Panchanathan, “Deep-learning systems for domain adaptation in computer vision: Learning transferable feature representations,” *IEEE Signal Processing Magazine*, vol. 34, pp. 117–129, Nov 2017.

[4] K. L. Masita, A. N. Hasan, and T. Shongwe, "Deep learning in object detection: a review," in *2020 International Conference on Artificial Intelligence, Big Data, Computing and Data Communication Systems (icABCD)*, pp. 1–11, 2020.

[5] Z. Ma, D. Chang, J. Xie, Y. Ding, S. Wen, X. Li, Z. Si, and J. Guo, "Fine-grained vehicle classification with channel max pooling modified cnns," *IEEE Transactions on Vehicular Technology*, vol. PP, pp. 1–1, 02 2019.

[6] M. Biglari, A. Soleimani, and H. Hassanpour, "A cascaded part-based system for fine-grained vehicle classification," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, pp. 273–283, Jan 2018.

[7] J. Sochor, J. Spanhel, and A. Herout, "Boxcars: Improving vehicle fine-grained recognition using 3d bounding boxes in traffic surveillance," *CoRR*, vol. abs/1703.00686, 2017.

[8] Q. Hu, H. Wang, T. Li, and C. Shen, "Deep cnns with spatially weighted pooling for fine-grained car recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, pp. 3147–3156, Nov 2017.

[9] F. C. Soon, H. Y. Khaw, J. H. Chuah, and J. Kanesan, "Pcanet-based convolutional neural network architecture for a vehicle model recognition system," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, pp. 749–759, Feb 2019.

[10] A. Nazemi, M. J. Shafiee, Z. Azimifar, and A. Wong, "Unsupervised feature learning toward a real-time vehicle make and model recognition," *CoRR*, vol. abs/1806.03028, 2018.

[11] L. Zhuo, L. Jiang, Z. Zhu, J. Li, J. Zhang, and H. Long, "Vehicle classification for large-scale traffic surveillance videos using convolutional neural networks," *Mach. Vision Appl.*, vol. 28, pp. 793–802, Oct. 2017.

[12] A. Dehghan, S. Z. Masood, G. Shu, and E. G. Ortiz, "View independent vehicle make, model and color recognition using convolutional neural network," *CoRR*, vol. abs/1702.01721, 2017.

[13] Z. Chen, C. Ying, C. Lin, S. Liu, and W. Li, "Multi-view vehicle type recognition with feedback-enhancement multi-branch cnns," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, pp. 2590–2599, Sep. 2019.

[14] C. Szegedy, S. Ioffe, and V. Vanhoucke, "Inception-v4, inception-resnet and the impact of residual connections on learning," *CoRR*, vol. abs/1602.07261, 2016.

[15] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[16] F. Bidoia, M. Sabatelli, A. Shantia, M. Wiering, and L. Schomaker, "A deep convolutional neural network for location recognition and geometry based information," 01 2018.

[17] M. Islam, M. F. Foysal, N. Neehal, E. Karim, and S. Hossain, "Inceptb: A cnn based classification approach for recognizing traditional bengali games," 05 2018.

[18] Y. Zheng, C. Yang, and A. Merkulov, *Breast cancer screening using convolutional neural network and follow-up digital mammography*. May 2018.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun 2016.

[20] M. Li, "Car image classification using deep neural networks," in *Colby College. Computer Science Dept, Honors Thesis (Open Access) Available at: https://www.semanticscholar.org/paper/Car-Image-Classification-Using-Deep-Neural-Networks-Li/68b69461f6943d6a6fd72c0176552b265f65789e*, 2019.

[21] M. Sheng, C. Liu, Q. Zhang, L. Lou, and Y. Zheng, "Vehicle detection and classification using convolutional neural networks," in *2018 IEEE 7th Data Driven Control and Learning Systems Conference (DDCLS)*, pp. 581–587, May 2018.

[22] D. Liu, "Monza: Image classification of vehicle make and model using convolutional neural networks and transfer learning.," in *Available at: http://cs231n.stanford.edu/reports/2015/pdfs/lediurfinal.pdf*, 2015.

[23] J. Yang, H. Cao, R. Wang, and L. Xue, "Fine-grained car recognition model based on semantic dcnn features fusion," *Journal of Computer-Aided Design & Computer Graphics*, vol. 31, p. 141, 01 2019.