Vehicle Classification using Deep Neural Networks and Transfer Learning for Surveillance and Theft Identification

Abhishek Angne, Glaston Dsouza, Mohammed Juned Shaikh
x18136923, x18132138, x18130224
Advanced Data Mining - Cohort B
MSc Data Analytics

Abstract— **The use of deep neural networks for the purpose of classification is one of the most powerful applications in the world of machine learning. With an overall progress in the world of technology, DCNNs have time and again proven to be the best classifiers, even in the field of image classification, thereby setting the benchmarks for numerous problems on classification. CNNs have been used for the purposes of classification back in the 80's. Vehicle identification is a critical application in the domain of traffic management, monitoring, and intelligent transport systems. This research aims at classifying 10 different classes of cars using D-CNNs from the VMMRdb dataset[1] using the Nvidia's graphics processor Intel's Kaby Lake processor using Tensorflow-Keras on the GPU in the Anaconda Environment. We have used transfer learning for the D-CNN models with the help of the ImageNet dataset as the base dataset and used it's learning to help the neural network perform better. The D-CNNs VGG16 and InceptionV3 models were used for performing multi-class image classification of vehicles. We could conclude that different models perform differently in terms of accuracy, precision, recall, f1-score, model size, training time and other evaluation metrics. In comparison to the traditional systems used for vehicle classification with the help of machine learning methods, D-CNNs have performed better.**

Keywords: Deep Neural Networks, transfer learning, computer vision, vehicle identification, TensorFlow Keras.

I. INTRODUCTION

The power of deep neural networks is reaching the zenith in a wide array of fields. Ubiquitous computing, security and surveillance, object detection, and classification. As per the recent trends, deep neural networks will work significantly in the fields of vocal user interface, machine compression, vocal UI, etc. These applications will increase the business value of deep neural networks in the fields of Artificial Intelligence.[2]

The idea of security and surveillance is the one that is one of the most powerful applications that are the driving forces of the domain. A television show named "Person of Interest"[3] was developed based on a system that could identify potential threats by accessing past data and facial and behavioral aspects of an individual and predict if the person will be harmed or will be a perpetrator. Apart from pictorial data, the algorithm portrayed was a kind of Voice-Activated Look-up algorithm which analyzed text and emotion based data. Technology giants like Google, Microsoft, Netflix, Facebook, and Pinterest are utilizing Deep Neural Networks at large in a multifaceted set of areas like Google's project called "Machine's That Dream" and Pinterest's algorithm to perform improved content discovery[4]. What we are going to perform in this research study, is transfer learning. To explain transfer learning in an analogous manner to human learning, think about it in this way, people have an innate capacity to move learning over various tasks or errands. What we learn as information while finding out whilst comprehending a single task, we apply that to interpret or solve tasks that are related to the initial task. For example,[5]if you have studied Statistics and Mathematics, you can use the knowledge to understand the fundamentals of machine learning task. Therefore, to attain a good classification we used the principle of transfer learning due to a smaller amount of training data for individual classes in the base dataset. With the help of transfer learning, we significantly sped up the overall process of training by utilizing aspects such as edge detection that the preceding training already knew.

A. *Research Question:*

"Can deep neural networks be used to perform significant multi-class classification and transfer learning on the Vehicle Make and Model Recognition Dataset (VMMRdb) of a chosen set of 10 classes of cars?"

B. *Research Objectives:*

The two deep neural networks chosen for this analysis are InceptionV3 and VGG16 from the family of Deep Convolutional Neural Networks. The research aims to determine whether the deep neural models perform multi-class classification on 10 classes of vehicles and perform significantly in terms of evaluation metrics like precision, recall, f1-score and accuracy.

Following are the objectives of the research that will that will test the model as per the given hypothesis:

[1]http://vmmrdb.cecsresearch.org/

[2]https://www.einfochips.com/blog/5-deep-learning-trends-that-will-rule-2019/

[3]www.imdb.com/title/tt1839578

[4]https:/www.wordstream.com/blog/ws/2017/07/28/machine-learning-applications

[5]https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

• Perform and visualize exploratory data analysis and fix the problem of class imbalance using Anaconda Prompt and Python.

• Classify ten different classes of vehicles using Convolutional Neural Networks(CNNs) using TensorFlow-Keras on the GPU.

• Identify the strengths and/or weaknesses of training the CNN model using transfer learning.

• Use hyper parameter tuning to improve the accuracy.

• Test the model based on various evaluation metrics like precision, recall, AUC curve and provide a performance evaluation of the model.

This paper is neatly structured in the following manner: II) Related work, III) Research Methodology, IV) Evaluation & Results and VI) Conclusion and Future Work

II. RELATED WORK

A. *Application of neural networks*

Yaxin Ma and team [1] proposed a new methodology using ResNet50 with a comparison with the VGG model in terms of accuracy, speed and other factors. The proposed method was fast and had no problems of over-fitting, thereby performed significantly. There was a vast difference between the training and validation set accuracy which eventually decreased with an increase in training time.

Xu [2] carried out a study of underwater image classification using deep CNNs using two augmentation techniques. The first was an optical transformation of images and color enhancement was performed. The second technique was to increase the virtue data generated using Generative Adversarial Networks (GANs). The GANs used in this study, consists of two sub models which are the generator (G) and the discriminator (D). The generator generates virtue images from scratch. The discriminator is a multilayer perceptron and it is trained to increase the probability to map correct label to virtual and data. The minibatch stochastic gradient descent was used to generate data in multiple categories. The GAN generated 3900 images targeting every class. Underwater captured images were tested for accuracy, which increased with the proposed augmentation methods.

In this paper, the author [3] classified various animals using ML techniques. Using Inception-v3, transfer learning was used to import the weights to increase the accuracy of the model. In the Inception model, features of previous layers are retained, and the final layer is replaced by the mammal dataset. The output nodes were changed to five since five different mammals were used for classification. The final layer was trained using back propagation algorithm and the cross-entropy function was used for adjusting the weight parameter by computing the error between the output of softmax layer and label vector of the image category. The result showed an overall accuracy 95% which was higher than the other methods.

In another research [4] an image classification method was proposed for crack detection based on deep fully convolutional network (FCN). Various CNN based algorithms were used in detecting cracks in roads, nuclear power plants and structures. In this study, VGG16, InceptionV3 and ResNet50 architectures were pretrained on ImageNet. A FCN network with VGG16 based encoder was trained on 500 labelled images for segmentation. The encoder retained pooling and convolutional layers and eliminated the softmax and FC layers of VGG16. The VGG16 with an Inception V3-based classifier performed the best and gave great results whereas the ResNet based classifier was less accurate.

In another study [5] transfer learning was implemented on Computer Vision Problems on the Plant Phenotyping database. The Google-Net CNN model was retrained on ImageNet dataset. A multi-label image segmentation was performed. The Inception model was trained on about 100,000 images and the dataset used in this study was comparatively small. Entire data was fine-tuned apart from the two top layers as they were pretrained and comprised of generic data-independent weights. The output layer was modified with a binary layer as only 3 types of plants needed to be classified. The results were highly accurate with 98% accuracy.

B. *Neural Networks for vehicle classification*

Fine grained classification of vehicles is a task of vehicle identification based on its manufacturer, model, color and year of launch. [6] compared performances of various CNN architectures which had shown great results in previous studies of vehicle identification (VGG16, ResNets, Inception models, MobileNet and DenseNet.) This study was conducted on Stanford Cars dataset. Different parameters of these architectures like augmentation and fine tuning were tested. The VGG networks used ImageNet for transfer learning. The results showed that ResNet-152 fine-tuned gives 92.6% accuracy and VGG16 gives 86.8%. The best result was achieved by DenseNet-161 model with an accuracy of 94.6%. As the dataset was not large enough, augmentation and fine-tuning the pretrained on ImageNet, gave better results.

According to [7], existing VMMR systems possess problems in multiclass classification tasks and face challenges like multiplicity and ambiguity in VMMR. The researchers have tackled these limitations using conventional systems by using VMMR with the ALPR systems. The researchers compared the outcomes of various network architectures and found that ResNet beats with the best accuracy. Hyperparameter tuning was performed and the baseline models showed better results for the images that were congested or under illuminated and had images taken at improper angles.

Baran [8] designed two different approaches in VMMR. The first method, Real-Time, is based on SURF detectors is used on real time recognition, where the priority is not given to accuracy. It was used at the training and testing phases. The second method, Visual Content, is designed to be highly accurate. The dataset consisted of 1360 images of 17 different models of cars. The SURF values are segmented using k-means clustering into specified number of clusters. The results showed that the RT approach gives 91.7% accuracy and VCC gives 97.2% accuracy. The author believes that these results can be incorporated with the grill of a car detector and can be used for car surveillance.

According to Li [9], the classification of cars is more challenging than general object detection since the similarity between some cars are very high. Also, the angle and illumination of the images plays a vital role in detection of cars. 3D object recognition car dataset was used as it comprised of different Sedans models with high complexity. The proposed algorithm first creates an edge map of an image. The results obtained show the segmented precision was 62.53% and recall was 79.09% when compared to manual segmentation.The proposed approach had satisfactory results and it would be interesting to apply it to larger and more complex datasets.

Liu [10] and team discussed about vehicle re-identification and proposed Deep Relative Distance Learning (DRDL) method using "VehicleID" dataset. The research is based on two datasets which are the VehicleID and CompCars dataset. DRDL minimizes the Euclidean distance between images of same vehicles and maximizes the distance between same vehicle images. Two very essential parameters which need to be considered in a re-identification model are; a technique for feature extraction and a distance metric which compares attributes of different images. A VGG CNN M 2048 model with its mixed difference was used as a feature extractor. The three evaluation methods used were vehicle model verification, re-identification and retrieval. In vehicle model verification GoogLeNet plus Joint Bayesian framework performed the best while the proposed DRDL approach stood second. In vehicle retrieval, it was found that the proposed method of coupled clusters loss function and mixed difference network structure helped in an increased precision as compared to other methods. For their research, the DRDL method performed the best. Vehicle classification method which used semi-supervised CNN was proposed by [11]. The dataset constructed for this study had 9850 high resolution front-view images of different vehicles known as BIT-Vehicle dataset. To obtain detailed features of the vehicles, Sparse Laplacian Filter Learning (SLFL) was used as it can capture the network filters of huge unlabeled data. The Softmax layer was trained on Multitask Learning with small chunks of labeled data. Huge amount of unlabeled data was required to train the filter bank of convolution layer in unsupervised manner. Whereas, the metrics of output layer are learnt with the labelled data in a supervised manner. The SLFL which is based on unsupervised learning, proposed in this paper, learns the filter bank of convolutional layer. The implemented method had an accuracy of 96.1% on daylight images and 89.4% accuracy for dimly lit images which was better than the outcomes of previous method. It was possible since the CNN used in this study was able to understand discriminative attributes of vehicle classification.

This study by Chen [12] proposed a system that could detect, track and classify vehicles using roadside CCTV footages. This study was conducted to perform statistical analysis on traffic data wherein a Gaussian Mixture Model (GMM) and a shadow removal method were used. A Kalman filter was used to identify and classify vehicles by majority voting through numerous consequent frames. The proposed GMM model was used for real-time tracking. This algorithm is based on the principal that the foreground appears less frequently than the background. Also, the model has comparatively low variance. A synthetic Measurement Based Feature (MBF, silhouette of the model) is established by projecting a vehicle wireframe model on the surface of the road to train the SVM model. Gaussian noise is introduced in the wireframe model, projected lane direction and position to create various kinds of vehicle silhouettes with noise effects in the image background. The best results are obtained from SVM which uses a blend of vehicle silhouette and intensity-based pyramid, HOG features. The results showed 96.39% detection rate. The accuracy obtained in unclear weather conditions was 94.69%. [13] discusses visual inspection in manufacturing plants and analysis of social media. A complete deep learning model which makes use of a mobile application for collection of data and Amazon cloud infrastructure for storage and training was used. AlexNet and GoogLeNet architectures were compared which were implemented on Caffe, Tensorflow and Torch. It was observed that Caffe 3 and Tensorflow 0.7.1 had better training times than Caffe 2 and Tensorflow 0.6 respectively. This can be due to the new cuDNN used in these frameworks. This can be due to the inconsistency in training times of epochs in the newer version. The cuDNN v4 and v5.1 were also compared wherein the newer version needed 25% lesser time than its previous version. Inception required the highest time for training amongst all. It can be stated that the more complex the architecture, the more time it takes for training. Tensorflow achieved the best peak accuracy of 94% with an average of 17 epochs. Also, the GoogLeNet required more training time compared to AlexNet but it had better scaling efficiency. For social media analysis, 106 car images were scraped from Twitter for model recognition. The standard results showed top-5 accuracy of 81.1% and F1 score of 85.9%. The region-search version gave slightly better results with top-5 accuracy of 82.1% and F1 score of 87.2%. The author suggested that distributed deep learning can reduce the training time for large and complex datasets. This research [14] addresses the car make and model identification considering within-category object recognition. The authors proposed an approach using a combination of global and local shape descriptors and using variation information labelled by human expert. Initially, the visually distinguishable parts from input images which corresponds to human-annotated parts were extracted. This correspondence is computed using global shape descriptors because the within-category classes had same global shape and part structures. After this, the extracted parts are defined with local shape and appearance descriptors. In the final step, the computed matching score of an image is evaluated by a weighted sum of global and local scores. Shape Context (SC) is used as a descriptor in global shape description to find correspondence between two points of different objects. The within-category images have similar global shapes and hence close SC descriptors which results in robust registration. The T2_8a_2r6s_PCA32 descriptor was used as it is fast and has low error rate. K-nearest neighbor (NN) with KD-tree indexing was used for feature matching. NN ratio helped in discarding outliers. The result obtained showed segmentation precision of 62.53% and recall of 79.09%. This method, therefore, enhances the classification rate where global and local descriptors are used.

In the research conducted by Zhou [15], a methodology was applied to analyze the vehicles based on the images and deep neural networks was applied to understand the patterns in the images and all the images were resized 448 in height and 333 in width which was achieved using the interpolation which means interjecting of two different data points into one. The neural networks used the weights of Alexnet and it was transferred to the neural networks. The latest algorithm called "You Only Look Once" YOLO which gave similar results but worked comparatively faster than the neural networks. In this analysis, it was found that the Deformable Parts Model performed the best with an accuracy and recall value of 93.3% and 83.3% respectively compared to the Dimensions based partitioning and merging algorithm which gave the results of 94.4% precision and 86.2% recall values. It was also noticed that out of the different layers of the model used the accuracy was significantly reduced when working with dark images but fc6 layer performed very well on dark images showing that it can handle the changes to the image in a better way images when compared to the fc7 layer of the model. In another research conducted on vehicle classification the authors inculcated the use of ensemble methods to classify the samples from the dataset. [16] used GANs with CNNs using semi-supervised learning. Data was collected from a traffic camera with high imbalance. Data was normalized using label smoothing. The ensemble methods proposed outperformed the deep learning algorithms by achieving better overall accuracy, recall and Kappa scores. Only used ResNets and Inception model weights were which was a limitation.

Michael and team [17] researched about a solitary class classification of VMMR. The authors used kNNs as their classifier and the analysis was done in Matlab. Their method involved edge detection. The classifier performed with an 97.5% accuracy with a drawback that it fails to recognize the exact model when the car model has internal changes but not external. The proposed model by [18] takes into consideration the various complexities faced during the image detection of the vehicles based on the size and capture of the image. The proposed methodology outperforms other traditional neural networks on basis of accuracy and speed. In this paper [19] has tried to solve the issue faced in terms of recognizing the make and model of a car from images as it is very difficult to analyze when the images are blurry. The analysis is done using the auto encoder of neural networks. This takes into consideration the edges of the data given to the system during the training period and the analyses gives the result on analyzing the different points on which the data is trained, eventually, performing with an accuracy of 94%.

III. RESEARCH METHODOLOGY

Methodology is one of the most essential bricks in the making of a successful research study. The research methodology selected for this research study is CRISP-DM.

[6]CRISP-DM was the methodology chosen because of its success as per the usage in multiple research studies conducted in the field of Computer Vision. CRISP-DM is the abbreviation for **CR**oss **I**ndustry **S**tandard **P**rocess for **D**ata **M**ining.

Fig. 1: CRISP-DM for Multi-Class Vehicle Identification

There are six different stages that the process of data mining is bifurcated into 6 subsections namely:

Stage 1: Business Understanding
Stage 2: Data Preparation
Stage 3: Data Preprocessing
Stage 4: Modeling
Stage 5: Evaluation
Stage 6: Deployment

A. *Stage 1: Business Understanding:*

The Vehicle Make and Model Recognition dataset also known as 'VMMRdb' is a dataset with a total of 9170 classes of different cars with over 2,70,000 images by Taffazoli and team [7]. Originally, the data was gathered by scraping data from the web from a number websites used for vehicle sales taken up from the Craigslist. The images are user-clicked images from various car-owners in the US from the years 1950-2016. For maintaining the privacy of the owners, the number plates were blacked out.

B. *Data Understanding:*

The car's selected for our analysis are the most stolen vehicles in the US as per an article published on the Forbes website and visualized by a data journalist named Mr. Niall McCarthy from Statista.com. A total of 10 classes namely, Honda Civic 1998, Honda Accord 1997, Ford f150 2006, Chevrolet Silverado 2004, Toyota Camry 2014, Nissan Altima 2014, Toyota Corolla 2013, Dodge Ram 2001, Gmc Sierra 2012, and Chevrolet Impala 2008 were chosen as per

[6]https://www.sv-europe.com/crisp-dm-methodology/

the survey of the vehicles that got stolen the most. As per an initial analysis performed by the authors, the 'Ford Explorer 2002' and the 'Nissan Altima 2005' were the cars stolen the most. Generally, sedans and pickup trucks were stolen the most.

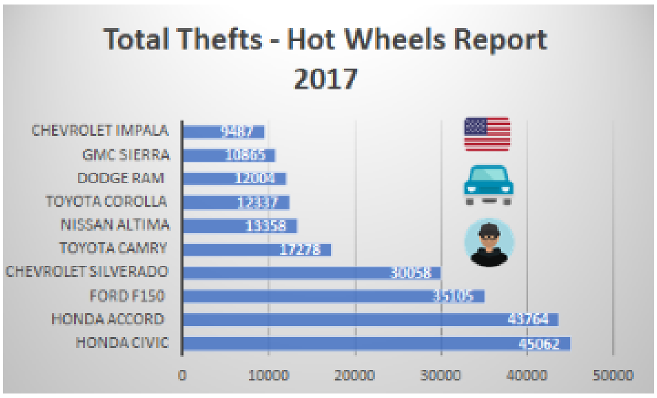

Fig. 2: Most Stolen Cars in the USA in 2017

The visualization shows the number of thefts in the year 2017. The most stolen vehicle in all the county's in the United States of America was the Honda Civic and the Honda Accord. Both these cars were Sedans which showed that Sedans were robbed most followed by Ford F150 which is pickup truck variant in automobiles. Out of the top 10, '6' most stolen cars were sedans and '4' were pickup trucks. SUV's and Hatchbacks were further down in the list of stolen cars in the US.

C. Data Preparation:

1) Choosing the right data:: The first step in data selection was to choose the required number of classes out of the 9172 classes chosen which is explained in stage 2. These classes were selected on the basis of visual parameters like the outer exterior. Let's say a 1997 Honda Civic and a 1998 Honda Civic had upgrades in the ECG(Engine, Clutch, Gear) section, however, the cars looked exactly identical, we have clubbed them under one class. Similarly, ten classes were created in terms of a visual criterion.

2) Removal of Corrupt Files:: Moving further, invalid files wherein the file was either corrupt or had some other issues, were removed with multithreading and the pool function. Two separate loops were utilized for removing the non-.jpg files.

3) Distribution of the 10 Classes Chosen for The Analysis:: The class distribution problem of our dataset existed in a few classes where the number of images present in the classes chosen were very less comparatively. To visualize the imbalance in the classes present in the subsetted classes of the total classes, we used pygal and galplot.

Figure 4 displays the imbalance in the chosen dataset

4) Intermediate Steps: The images were re sized to a width and height of the value 299 each. The intermediate steps include the confirmation of folder structure of the 'Most_stolen_cars' to view the images used. This was followed by creating the train, validation and test folders using

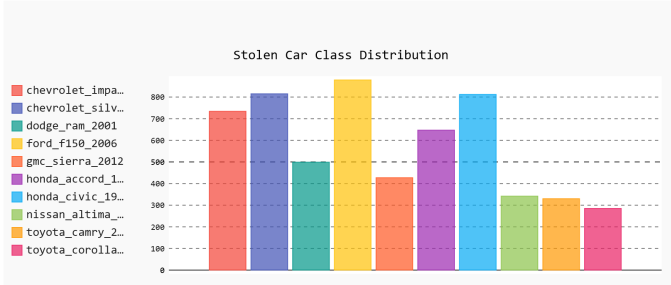

Fig. 3: Class Imbalance

85%, 5%, and 15% respectively. Augmentation was carried out on the images at a rotation of 45 degrees, along with vertical and horizontal flipping.

5) Fixing the class balance: The class distribution was checked once again after augmentation using flipping and rotation.

The following image displays the class balanced better than the initial position of the subset classes.

Therefore, the classes with lower samples were up sampled with the help of augmentation techniques. Below Figure 4 shows the classes after balancing.

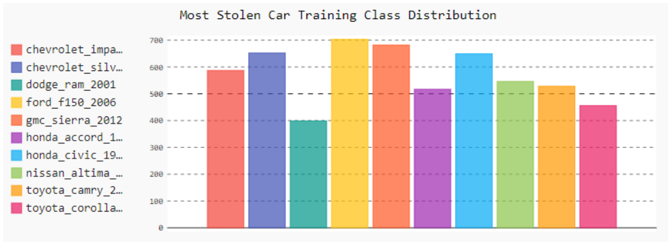

Fig. 4: Class Balance

6) Training Model:: For fraud classification models, we have applied the 10 fold cross validation and split the training and testing data in 80:20 ratio.

D. Modelling:

1) VGG16: The Visual Geometry Group from Oxford developed their own Convolutional Neural Network called VGG16, which is also called as OxfordNet.

VGG16 is a deep convolutional neural network (D-CNN) that was proposed by Karen Simonyan and Andrew Zisserman in the paper Simonyan2014

This CNN achieves a top-5 testing accuracy of 92.75% for the ImageNet dataset. The ImageNet dataset comprises of about 14,197,122 number of images belonging to a total of 1000 classes. This model takes an input image of the dimensions 224 as the width, 224 as the height, and 3 for RGB pixels corresponding to Red, Green and Blue colors.

The research on large-scale recognition of images with the help of Deep Convolutional Neural Networks(CNNs) that concluded by highlighting the importance of representation depth in terms of accuracy of the classification.

Following is the architecture of VGG16

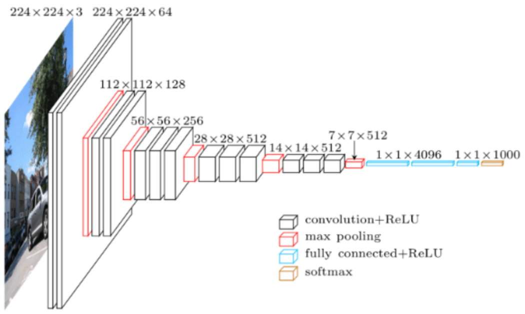

Fig. 5: VGG16 Architecture
7

Hyper Parameter Optimization

In the initial stages, fine tuning was done on only the top layer. For optimization, all the layers in the network to be trained were tuned using a lower learning rate. Therefore, the network could tune the weights we hadn't tuned in the first place from the checkpoint of ImageNet.

The freezed layers were unfreezed and the model was recompiled again. The remaining part of the code was the same except they now indicated that we were using the upper 2 inception nodes during training.

For this research in machine learning, classification deep learning algorithms were evaluated using F1 score, , Precision, Recall, and Accuracy, were the metrics considered for the analysis.

2) InceptionV3: InceptionV3 has shown an accuracy of 78.1 and more on the famous ImageNet dataset. The Inception V3 model is a pre-trained model on the ImageNet Dataset which is a systematically organized according to a hierarchy called "Wordnet" wherein every noun is allocated with thousands of images. The model after learning from this dataset uses the knowledge of the ImageNet dataset and applies it for the processing of a new data set. Moving on to multi-class classification which is the type of classification we're aiming to achieve at the end. Multi-class classification works on the assumption wherein every sample gets assigned with a unique label, for instance a fruit can be an orange or a mango but cannot be both the fruits at the same instance.[8] Out of the 22 layers of the inception model, a few layers decode features like blob or edge detectionl. Layers after these, perform shape detection and color detection. In the final layers more complex attributes pertaining to specific categories are captured. In this manner graphical attributes of fed images are learned by the model.

In the research conducted by Gavai2018 on Image Classification, Inception V3 was used to classify different classes of objects. Classifying various objects as per their respective classes has been a very significant application in the world of machine learning. The aforementioned research Classification of objects into their specific classes is always been significant tasks of machine learning.

[8]https://towardsdatascience.com/machine-learning-multiclass-classification-with-imbalanced-data-set-29f6a177c1a

The research tries to classify various species of flora using computer vision. In this research, the team found that Inception V3 took more time comparatively. This is the reason we tried to compare the performance of the Inception V3 model with the VGG 16 model.

Following is the architecture of InceptionV3:

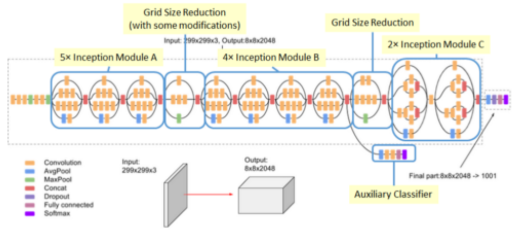

Fig. 6: InceptionV3Architecture
9

IV. EVALUATION OF RESULTS

1) InceptionV3: The Inception V3 model was evaluated on a batch size of 32 and a loop of epochs of 170 and the model was optimized using freezing methods. The model performed relatively bad on the amount data fed to the model for training. The model gave an overall accuracy of 58%. The model also failed to identify many cars and gave a lot of poor results for both recall and precision.

2) VGG16: The VGG 16 model was similarly trained with same amount of batch size and epochs using similar optimization techniques for the VGG model. The model gave very good results and gave an overall accuracy of 82%. The model identified many classes accurately and gave high values for both precision and recall. The overall model performed really well but it executed the iteration of code at a relatively slower pace and that was the only drawback during the execution of this process.

Precision: A higher value of precision translates to a lower rate of values that are false positives. Precision is simply a ratio of true positives over the total number of positives.

Precision tries to signify the actual positives out of the positives predicted. Let's say a data on survival was being analyzed, precision tells you the actual number of survivors out of the ones that were labeled or termed as survived.

$$precision = \frac{TP}{TP + FP} \tag{1}$$

Recall: Recall is also called as sensitivity. Mathematically, recall is a ratio of the true positives or TP to the sum of false negatives(FN) and true positives (TP).

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

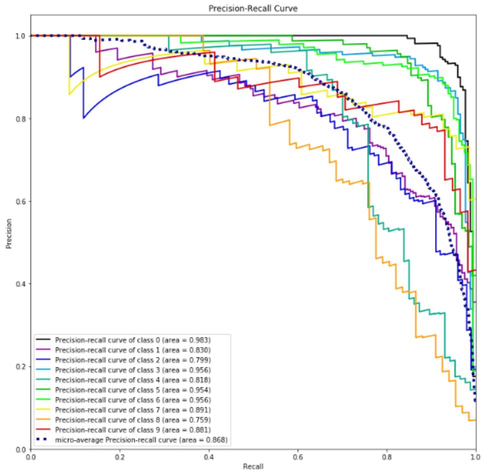

Fig. 7: P-R Curve

f1- score: The f1-score is the harmonic mean of precision and recall values. It's simply the weighted average of precision and recall.

$$f1score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

Receiver Operating Characteristic Curve(ROC Curve): The performance metric named AUC - ROC curve is used for the measurement for issues related to classification. ROC is more of a probabilistic curve whereas AUC depicts the degree of separability. It helps in understanding the capability of the model for differentiating between various classes. More the value of AUC, better is the performance of the model at predicting a value. A good prediction is when a 0s are predicted as 0s and 1s are predicted as 1s.

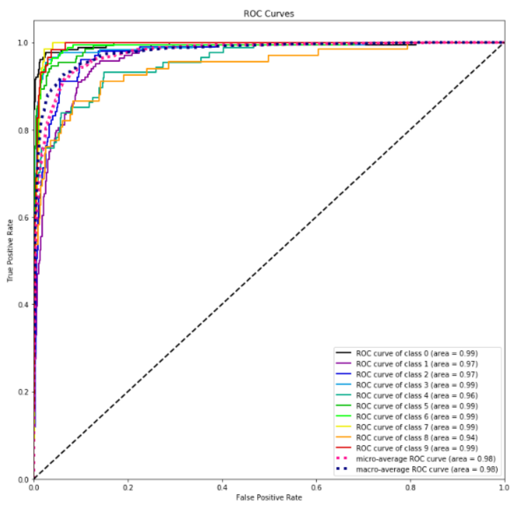

Fig. 8: ROC Curve

```
In [76]: from sklearn.metrics import classification_report
         report = classification_report(
             true_classes,
             predicted_classes,
             target_names=class_labels)
         print(report)
```

	precision	recall	f1-score	support
chevrolet_impala_2008	0.95	0.99	0.97	176
chevrolet_silverado_2004	0.81	0.68	0.74	164
dodge_ram_2001	0.84	0.55	0.67	101
ford_f150_2006	0.67	0.97	0.79	177
gmc_sierra_2012	0.83	0.63	0.72	87
honda_accord_1997	0.75	0.94	0.84	131
honda_civic_1998	0.96	0.84	0.90	163
nissan_altima_2014	0.75	0.88	0.81	69
toyota_camry_2014	0.93	0.55	0.69	67
toyota_corolla_2013	0.85	0.81	0.83	58
accuracy			0.82	1193
macro avg	0.83	0.79	0.80	1193
weighted avg	0.83	0.82	0.81	1193

Fig. 9: Accuracy

Confusion Matrix: The confusion matrix signifies the prediction performance thus analyzing the network. The

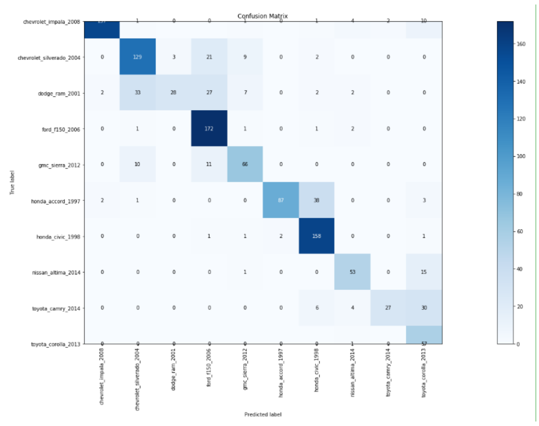

Fig. 10: Confusion Matrix

V. CONCLUSION AND FUTURE WORK

A widely used dataset has been used to implement different neural networks on the data as well as to analyze the results to know how it has performed on the given dataset. In the proposed research, we have used the VGG16 model to correctly predict the make, model and year of the car from the subset of the dataset. We have implemented the VGG16 model by importing weights from the imagenet model in keras which is pre-trained on a seperate data. The results show that even though the model is not able to outperform the state of the art frameworks it can be improved by using larger sets of data as well as implementing more augmentation techniques. The model performed really well and gave an overall accuracy of 82% which can be improved by training the data on more images as well as by reducing the batch size and increasing the number of epochs. The model however performs slow in terms of speed and in future work other models will be used using a larger dataset to increase the speed and performance of the deep neural network.

References

[1] Y. Ma, P. Zhang, and Y. Tang, “Research on Fish Image Classification Based on Transfer Learning and Convolutional Neural Network Model,” *2018 14th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pp. 850–855, 2018.

[2] Y. Xu and H. Wang, “Underwater Image Classification Using Deep Convolutional Neural Networks and Data Augmentation,” 2017.

[3] J. Bankar and N. R. Gavai, “Convolutional Neural Network based Inception v3 Model for Animal Classification,” pp. 142–146, 2018.

[4] C. Vu and L. Duc, “Automation in Construction Autonomous concrete crack detection using deep fully convolutional neural network,” *Automation in Construction*, vol. 99, no. July 2018, pp. 52–58, 2019.

[5] A. Tapas, “Transfer Learning for Image Classification and Plant Phenotyping,” vol. 5, no. 11, pp. 2664–2669, 2016.

[6] K. Valev, A. Schumann, L. Sommer, and T. Kit, “A Systematic Evaluation of Recent Deep Learning Architectures for Fine-Grained Vehicle Classification,”

[7] F. Tafazzoli and H. Frigui, “A Large and Diverse Dataset for Improved Vehicle Make and Model Recognition,”

[8] R. Baran and A. Glowacz, “The efficient real- and non-real-time make and model recognition of cars,” pp. 4269–4288, 2015.

[9] M. Li, “Car Image Classification Using Deep Neural Networks,” 2019.

[10] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang, “Deep Relative Distance Learning : Tell the Difference Between Similar Vehicles,”

[11] Z. Dong, Y. Wu, M. Pei, and Y. Jia, “Vehicle Type Classification Using a Semisupervised Convolutional Neural Network,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 2247–2256, 2015.

[12] Z. Chen, T. Ellis, and S. A. V. Smieee, “Vehicle Detection , Tracking and Classification in Urban Traffic,” *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pp. 951–956, 2012.

[13] A. Luckow, M. Cook, N. Ashcraft, E. Weill, E. Djerekarov, and B. Vorster, “Deep Learning in the Automotive Industry : Applications and Tools,” 2017.

[14] M. Abdelmaseeh, I. Badreldin, M. F. Abdelkader, and M. E. Saban, “Car Make and Model recognition combining global and local cues,” *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, no. Icpr, pp. 910–913, 2012.

[15] Y. Zhou, H. Nejati, T.-t. Do, N.-m. Cheung, and L. Cheah, “Image-based Vehicle Analysis using Deep Neural Network : A Systematic Study,”

[16] W. Liu and Z. Luo, “Knowledge-Based Systems Improving deep ensemble vehicle classi fi cation by using selected adversarial samples,” *Knowledge-Based Systems*, vol. 160, no. 422, pp. 167–175, 2018.

[17] G. Michael, T. Daniel, D. T. Munroe, and M. G. Madden, “Multi-Class and Single-Class Classification Approaches to Vehicle Model Recognition from Images Author (s) ” Multi-Class and Single-Class Classification Approaches to Vehicle Model Recognition from Images ”, Daniel Munroe and Information Multi-Class and Single-Class Classification Approaches to Vehicle Model Recognition from Images,” 2005.

[18] H. Tayara, K. Gil Soo, and K. T. Chong, “Vehicle detection and counting in high-resolution aerial images using convolutional regression neural network,” *IEEE Access*, vol. 6, pp. 2220–2230, 2018.

[19] Y. Yang, Y. Lai, G. Zhang, and L. Lin, “A vehicle recognition method based on adaptive segmentation,” in *Communication Technologies for Vehicles* (J. Moreno García-Loygorri, A. Pérez-Yuste, C. Briso, M. Berbineau, A. Pirovano, and J. Mendizábal, eds.), (Cham), pp. 125–136, Springer International Publishing, 2018.

Appendix

Team Contribution:

Name	Contribution
Abhishek Angne	34%
Glaston Dsouza	34%
Mohammed Juned Shaikh	32%