IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12324

Vehicle Classification using Deep Learning

Mohammed Afnan Ahmed[1], Manideep[2], Mohammed Zaid[3], Mohammed Sudais Khan[4], Prof. Janhavi V[5]

UG student, Dept. of CSE, Vidyavardhaka College of Engineering Mysore, India[1-4]

Associate Professor, Dept. of CSE, Vidyavardhaka College of Engineering Mysore, India[5]

Abstract: Vehicle Classification has wide applications in intelligent transportation and smart cities. Classification is crucial for an intelligent transportation system (ITS). Vehicle make and model classification technique is very useful. Make and model is a fine-grained information that can help officers uncover cases of traffic violations when license plate information cannot be obtained. Faster-RCNN and YOLO are two algorithms which are used for Object Detection and Classification. These algorithms work very well for classifying different types of vehicles. The various vehicles are classified into different vehicle classes. Initially, the dataset is made ready using the videos or converted images format. The dataset needs to contain the data which is of the predefined classes types. Secondly, the images or videos not used during the training are then used for the evaluation of vehicle classifier model. The main parameters that are used in the evaluation of the model include training time, the testing time, the vehicle classification accuracy, as well as the performance of the specific deep learning methods. Many of the datasets are used in the classification of datasets, it makes the evaluation of the classes of datasets more simpler and easier. The previous data could be pre-processed to generate the new datasets that could be used for the evaluation of datasets.

Keywords: Vehicle classification; color classification; deep learning; convolutional neural network; Faster R-CNN, YOLO, VGG-16.

I. INTRODUCTION

Deep Learning in contrast with machine learning, which is a domain where the machine is trained with minimal human interference, is a subset of machine learning that uses artificial neural networks to mimic the learning process of human brain. It has been widely used in medicine [1], education [2], mining [3], and transportation [4]. In every ANN, there are 3 layers of neurons or nodes, namely, input layer, hidden layer and output layer. The weights are assigned to each of the connections made, these weights and bias factors could be altered during the process. The final weight values are considered to be the most accurate values as of that moment. Due to these weight values, the new image get classified best according to the already existing classes. In deep learning, a convolution neural network (CNN or ConvNet) is a class of artificial neural network (ANN), applied for analyzing visual imagery. They have applications in image and video recognition. Classifying the objects found in the images and videos. The object detection is the first step that takes place when CNN is used. Object detection is a technology that detects the objects based on the features of it. Firstly, all the objects that exist in the image are identified and then the objects are filtered. It has various applications like face recognition, vehicle recognition, pedestrian counting, self-driving vehicles, security systems, and a lot more. Some of the most used Object Detection techniques are as follows: ImageAI, Single Shot Detectors, YOLO (You Only Look Once) and Region-based Convolutional Neural Networks. Classification is done using pre-trained models. Some of the most commonly used models are VGG-16, ResNet50, Inceptionv3, EfficientNet. These are the top-most models that are used for classification. The VGG-16 is one of the most popular pre-trained models for image classification. It has the following layers of the model, Convolutional Layers = 13, Pooling Layers = 5 and Dense Layers = 3. Inception Model performs convolutions with different filter sizes on the input, performs Max Pooling, and concatenates the result for the next Inception module. ResNet50 model aimed mainly to tackle the issue of Vanishing Gradient Descent. It avoided poor accuracy as the model became deeper. EfficientNet uses a new scaling method called Compound Scaling. The earlier models like ResNet follow the conventional approach of scaling the dimensions randomly and by adding up more and more layers. The scaling, if done by a fixed amount at the same time, then much better performance is expected. The scaling coefficients can be decided by the user. By using both Object detection and classification algorithms, the vehicles could be classified into different categories.

Vehicle classification (VC) is an underlying approach in an intelligent transportation system and is widely used in various applications like the monitoring of traffic flow, automated parking systems, and security enforcement. Vehicle detection and vehicle classification using neural network (NN), can be achieved by video monitoring systems. In most vehicle detection methods, only the detection of vehicles in the frames of given video is emphasized. To obtain more information regarding traffic management and number of vehicle types passing through the roads, more analysis is needed. The

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12324

accuracy of vehicle detection and classification is vital for further ighway transport analysis. A wide variety of information could be gathered and extracted using sensors and detectors which may include vehicle count, shape, speed, acceleration/deceleration, make and model [5] and number plate.

This paper covers the following:
i) Related Work
ii) Proposed Methods
iii) Evaluating Classification Models
iv) Applications of Classification
v) Conclusion and References

II. RELATED WORK

There were many studies and researches carried out for the classification task. Some of the most famous and most used algorithms are neural network, logistic regression and SVM for classification. The comparison has been done by D.K.R Hosteller with a dataset of 3074 samples. Experimental results have shown that logistic regression has given the better performance. If the dataset consists of various size and shapes of vehicles then SVM and Random Forest Classifier is the best algorithms to choose from. The SVM performed better in comparison with Random Forest Classifier in the classification of car and van. M. Mazoor has proposed a system of linear SVM for the classification task that provides more accurate classification as compared to the other algorithms that do the same classification task. The Scale Invariant Feature Transform is used to extract features and local regression analysis. Some of the procedures were carried out which had the background subtraction, foreground differentiation and features getting extracted. SVM was the best performer during day time but the results were not up to the mark during night time and bad weather conditions.

Many statistical methods are used such as Support Vector Machines (SVM), Efficient Support Vector Machines (eSVM), Bayesian Discriminant Features (BDF), Adaboost, clustering-based discriminant analysis. Some of the other algorithms used were Scale Invariant Feature Transform (SIFT) [6], Feature Local Binary Patterns (FLBP) and Histogram of Oriented Gradient (HOG). Fig 1. shows the overview of the system, the video features are extracted. The relevant features are the only ones that are needed to be kept, so the remaining ones are discarded via pre-processing step. Further classification and selection is done via the Filtering step. Filtering step is a step where some filters are applied to the existing dataset, and the dataset values or the things that pass through the filter or in other words get through the passing criteria of the filter are then considered for further steps, rest of the data values and the data points are discarded. These relevant features are then gone through a process of classification where the data set points are analyzed. They form into two different categories, one set of points that are similar in nature and the rest of the points that are dissimilar in nature. The classification of data points does the exact same thing. Now, since the model is trained, it is kept in database for further usage and classification purpose. The search manager is used to search in the database any relevant thing which is close to what the system query from the user side is expecting. This query then gets processed and further elaboration in the semantics and nuances of the query fetches some relevant results. Deep Learning methods and Neural Network methods are used. Among the many Deep Neural Networks (DNN), Convolutional Neural Network (CNN) is one of the architecture. In every neural network there are three layers, input, hidden and output.Although, the Deep Neural Network is used to train the model and classify the images, CNN works on a large data input with useful patterns and similarities that are interrelated and interconnected with each other. Iteratively, the CNN extracts the relevant features and thereby updates the weights and outputs the probable class with the accuracy. The error is calculated by the subtracted difference between actual output and expected output. These error values which are the delta error or the subtracted difference then forms the factor for gradient error term. This term is used to add and subtract the weights so that it can be made the best possible values. These new values are then again tested iteratively for further changes in the weights and in search of getting the correct output.

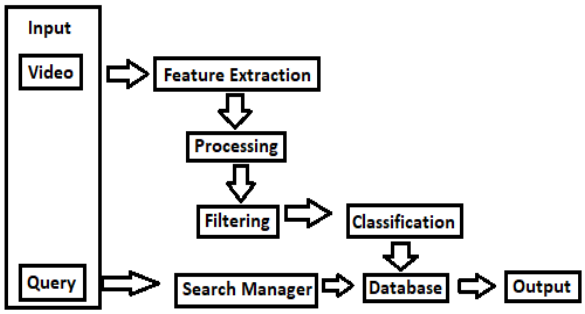

Fig 1. System Overview

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023
DOI: 10.17148/IJARCCE.2023.12324

CNN extracts features using convolution filters and pools like max-pooling, then finally the activation layer filters such as ReLu (Rectified Linear Unit). The Region based CNN (R-CNN) uses region proposals through selective search that applies different sizes of window to evaluate the entire image dataset. Many Regions are extracted in the selective search process, and it applies a custom version of AlexNet, which determines the valid region. Another very efficient algorithm, is Fast R-CNN which performs better than R-CNN. R-CNN becomes slow because of the need of forward pass for each proposed region individually. The Fast R-CNN is a really fast algorithm even faster than R-CNN because of combining different parts of the process and sharing computations. The first step is to feed the entire dataset to the CNN.

The region proposals generate image share and maps the features along with the network layers, thus the speed of the model increases by reducing the time needed to accumulate the content. Faster R-CNN eliminates the proposal of region and instead uses another network called Region Proposal Network. It generates object proposals and also learns from the Fast R-CNN network. You Only Look Once (YOLO) was introduced in 2015. It is a deep learning method that divides the image input into S x S grid, and only one object detection happens in one cell, and the center of the object falls into the cell when the detection happens. In each cell, a fixed number of bounding boxes with their confidence scores are generated. The confidence scores are calculated by multiplying the probabilities of each object and their union is calculated after intersection.

Further, YOLOv2 and YOLOv3 were introduced with the overcoming of the limitations that the previous versions had. Finally, it is claimed that YOLOv3 is three times faster and has even similar accuracy.

III. PROPOSED METHODS

a. Categorization of methods

Classification of vehicles or in general image classification could be achieved through various methods. It depends on the situation and the dataset that is being considered for the evaluation purpose, that which method would be the most suitable. Deep Learning methods are used in classifying the image datasets. The most widely used method is the implementation of Convolution Neural Network (CNN). Convolution Neural Network could be used to evaluate the datasets for setting the weights and classifying the output values. One of the most widely used method of carrying out this task is the implementation of Transfer Learning technique. Transfer Learning technique is the method that uses pre-trained models for the classification of object entities. This method takes in the input parameters in the form of image datasets. These image datasets are then further classified as various output categories. Image classification is one of the domains where the concept of deep learning excels.

The dataset of images is provided, it is then classified into various categories based on the weights assigned in the pre-trained model. Transfer learning is a super method that could provide the classification results with very less effort spent towards the building model from scratch. When things are built from scratch it takes a lot of time, but when things are ready and they just have to be implemented it takes very less time. This is because building something from scratch requires that we know the requirement for something, the content required for it, the methodology for the implementation. The output depends on the convolution layers used in between, the max pooling and the activation functions such as soft max or relu used in the execution of the model. A very big factor that effects the results evaluation is the fact that the dataset also plays a very vital role in the classification of concepts. When something is already trained and we just have to use it, it becomes so much more easier than actually implementing from the scratch. Same goes with the pre-trained models that are used in transfer learning. The pre-trained models help us evaluate many things in a much less time because it is already implemented. Computer Vision is the tool we would be most frequently using for the evaluation of the datasets. Computer Vision is the method using which objects are detected. They have the features to identify the objects present and then classify them on the basis of the object features. This method of classification is known as computer vision image classification. Computer Vison gathers many different domains under a single domain. The various domains are machine learning, deep learning and artificial neural networks for classifying the images.

Method 1

The first method of classification of images is the method of using transfer learning. It removes the final layer of Inception which is v3 layer. Using this, it performs the process in a more better and optimal manner. For the feature vector of an image to be classified there has to be a retention of the last layer in the network, this enhances the classification of the layer. Few of the algorithms are used in achieving it are support vector machine, K-nearest neighbours [7] and deep learning. The training data is passed to the inception v3 model which is used as a transfer learned technique. This enhances the output got from the retention of the final layer. The model fine-tunes the required parameters and makes sure that all the layers are present as a extension of the existing layer. Tensorflow performs the classification by using the flow of images. It converts the image dataset into the feature vector and classifies the images accordingly. The dataset used was

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023
DOI: 10.17148/IJARCCE.2023.12324

VeRi which had 50,000 images with over 700 types of vehicles and 20 camera views from which the dataset was prepared. The images of the same instance from 20 cameras are taken as a single unit and provided to the computer system as the single feature vector value taken from different dimensions and angles. The ratio of train test set was 2:3 and the results obtained were significant. KNN had performed very well on the VeRi dataset with an accuracy of 97%. SVM performed the classification with an accuracy of 90%. The conclusion is that the KNN performed better than SVM.

Method 2

This method used the concept of multi-task learning and classifies the images based on the CNN approach. For the better output and optimal approach, the VGG-16 architecture was taken. Most of the vehicles are classified on the basis of their features, this has to be made clear that the feature is nothing but mainly consist of the make and model of the vehicle. Histogram of Gradient [8] is also a approach for vehicle make and model classification. There are situations where the license plate information cannot be obtained. This case the police officers might find it difficult to make the identification of the vehicle that is being taken into consideration, but if a model is built on the basis of the brand and the manufacturer, then it becomes much more easier for the police officer to detect the vehicle. This type of implementation is done in the following method, for more accurate results the architecture used in CNN is the VGG-16 model. The model makes two ways and then it combines the outputs of both to classify the vehicle. The model does this parallelly, it makes one way to classify the model and the other way to classify the make.

These outputs are then combined to get the final output. This model is mainly designed for highly similar models, therefore for the implementation of the following, the indonesian dataset is used InaV-Dash dataset is used which has the models and makes of vehicles which are very similar in nature. The experiment shows the proposed method achieved for the vehicle make an accuracy of 98.73% and achieved an accuracy of 97.69% for vehicle model. The proposed method also performs well on the dataset other than the given dataset and classifies very well the highly similar datasets. The functions used and the applications of the thing that processes the things is also used. The convolution neural network architecture used the max pooling approach for the better approach of the dataset. The approaches are explained in great depth. At the end the methods of evaluation of the performance of the model is also explained in great depth. Single task and Multi task approaches are used for the implementation of the model. Single task approach uses the single classification for a specific task and other layers classification for other attributes or features. Single task layering requires just one attribute whereas multi task layering can make use of multiple attributes and then if further layering is needed it just has to change the final layer to get the best result.

It is very much time efficient. The proposed multilayer neural network consist of Max-Pool Layer, Global Average Pooling Layer, Dense Layer, Dropout Layer, Activation Functions, Loss Function. The performance evaluation model is evaluated on the basis of the confusion matrix. Confusion matrix takes into consideration false and true values of the data set points. The summarization is done via the macro average technique. Keras Library which is connected with the TensorFlow backbone was used to construct all the CNN layers. ImageNet was used as a pre-trained model. The result is that ResNet was able to reach the convergence after 16 epochs, whereas VGG models were producing higher errors during training errors as compared to other architectures. The measurement was performed using a InaV-Dash dataset. Accuracy obtained by ResNet50, Inception, InceptionResNet, and MobileNet, was 93%. VGG-16 and VGG-19 obtained 56% accuracy. Future scope includes, increasing the dataset and adding the video data for the evaluation as well. This sheds light on the next tasks and challenges in the domain of CNN.

Method 3

This method uses the CNN architecture and neural network model that classifies the images based on the model and the color of the vehicle. K. Ying et al. [9] proposed a decision tree as a classifier. The input is just a single image data and the classification is done on the basis of the features of the vehicle. The proposed CNN consists of the following attributes and their values:

Input consists of a 32x32 resized image, the convolution layers are present at the 1st and 2nd layer, the pooling layers are 2, Activation functions used are 4 ReLU functions, 3rd, 4th and 5th layer are fully connected layer and the dropout layer is 1. The output consists of number of possible layers as shown in TABLE I.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12324

TABLE I. COMPONENTS OF THE PROPOSED CNN

Input	A 32x32 resized image
Convolution layers	1st layer and 2nd layer
Pooling layers	2 pooling layers
Activation functions	4 ReLU functions
Fully connected layers	3rd layer, 4th layer and 5th layer
Dropout	1 dropout
Output	The no. of possible classes

The proposed method is evaluated via three schemes datasets, environments, and evaluation results. The dataset consists of video duration taken as 10 min long video, resolution is 640x480 pixels, frame rate is 25 fps. extracted vehicle images are above 900. The ration of train test data is 3:1. The environment used is jupyter notebook along with the TensorFlow. Softmax is the classifier, dropout is 40%. The Adam Optimizer is initialized as 0.001. Batch size at the start of the no. of epochs is 5,000.
The evaluation result thus obtained depicts the fact that CNN classified the test images with and accuracy of 85% in type and 74% in color classification. For future scope, AlexNet is used which classifies the dataset images with more accuracy.

Method 4
In this method, the main used techniques are the Faster R-CNN and YOLO v3 methods. In Faster R-CNN method, it uses the VGG-16 architecture. It is a very powerful convolution neural network architecture that could be used for the calculation of output, given the input dataset. The RPN (Region Proposed Network) is used extensively throughout the network for the better region proposals. Local Patterns such as Local Binary Patterns (LBP) is used [10]. Fine-tuned samples are used for the better training and testing, hence the evaluation becomes better over time. Window slides are used for the entire feature map, at the center of each sliding window there are anchors that point to the three different scales and three different vectors are generated. VGG-16 has 13 convolution layers, 5 pooling layers, 3 fully connected layers and 1 softmax classification layer. The size of the convolution filter was 3x3 and size of max pooling layers was 2x2. The down sampling is done to produce more robust features. Fully connected layers are responsible for the calculation of each score and to the generation of output. The fully connected layers are adjusted for different task visions. YOLO v3 had not achieved the real-time performance. So it was improved by a method proposed in the paper. The method mainly had an accuracy increase over time and the real time performance of the data by improving the version of 13 layers instead of 65 layers. For the consideration purpose there are 6 max pool layers used. Each of the cell in the input grid GxG, there is a part of the dataset taken into consideration. Bounding box consists of 4 parameters, quadruples. The parameters are the following: (x, y, w, h) are the four parameters used to determine the box where the image has to be detected, the x and y coordinates are the coordinates of the bottom left corner. The remaining coordinates are found using the width and height calculation. False positives and False negatives are used for the calculation of error rates. Identification of the corresponding vehicles, the frames are used in dataset evaluation of the results. The frame processing step is YOLO which is 30 times more quicker than the Faster R-CNN. It is due to this feature of the YOLO that it makes the selection of real-time dataset. Latest versions of YOLO have better and improved accuracies.

Method 5
In this method there are two types of classifiers used, Single Classifier and Multilevel Classifier. In single level classifier it is observed that the ending layers are trained first. When convolution neural layers were released the training was done on the entire network [11]. The lower layers were trained with a lower accuracy. There are mainly two datasets to be taken into consideration, BIT dataset and Stanford dataset. In BIT dataset, the main used architecture was the ResNet34 which used the different and varying sizes of the images. The different sets of images were taken into consideration. The documentation of the result of training is done, so that the accuracies and the errors were taken the note of. The image size varies inversely with the training and validation loss. The accuracy increases from 44% to 77% when there was an increase in image size from 32 to 224. The time spent increases from 27 to 81 minutes. Trainings on various datasets were done on size such as 32x32, 64x64, 128x128 image sizes were used. 150x150 was used for the ResNet50 and 224x224 was used for ResNet34. In stanford dataset, single level hierarchy is the most widely used approach for training

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12324

models. ResNet34 uses varying image sizes. The observation was that the accuracy increases with the image size, but along with that the training time also increases. Accuracy increased by increasing the image size. Along with ResNet34, ResNet50, ResNext50, VGG-16 architectures were also used. Accuracy for BIT dataset is 44% and Standford Dataset is 28%. The future scope includes the following of 3 mechanisms. The better results are obtained if the dataset is enlarged and studies are repeated with large image dataset. Pre processing can be done before training a model. The vehicle could be cropped and studied with focus. The background image is the noise data. The success relatively means more multi-level training, no time issue produces more results.

b. Vehicle Classification using VGG-16

To utilize VGG-16 for the best performance in classification, the two types of learning are taken into consideration. Single task and Multi-task learning. The modifications that are made in terms of type of layer, no. of neurons, activation function are explained in great depth. Single task is used when the attribute is single and the classification is needed to be done for the same. For training the model for a new attribute the model has to be re trained from the beginning then it causes a lot of time and resources. The alternative approach that is better than single task learning is the multi task learning. In multi task learning, the task learns from a perspective from a multiple attributes. It saves a lot of time and resources. The visual representation of single and multi task learning is given in the Fig 2. (Single Task Learning) and Fig 3. (Multi Task Learning). In Single Task Learning, the attributes and the data input in the form of x1, x2, … , xn are given. These are passed to the Hidden Layer which performs all the operations and the procedures that are needed for the classification. The normalization is done at the output layer. Similarly, in the Multi Task Learning the data inputs remains the same whereas the inputs that are passed into the Hidden Layer then get divided into multiple specific task based outputs. Multi-Task Learning mainly classifies the 4 types of vehicle brand classes and 10 models. They are very much closely related to each other. CNN is a robust method that uses the artificial neural network concept to classify the input to different output classes. It updates the weights according to the error that is obtained after each epoch. Most of the vehicles are classified on the basis of their features, this has to be made clear that the feature is nothing but mainly consist of the make and model of the vehicle. Multi-task learning helps save time and make the efficiency more better. The project fits well with the Multi-task learning approach. It classifies multiple classes of objects in the input. The vehicle make and model classification uses Multi-Task Learning approach. So the classification is needed

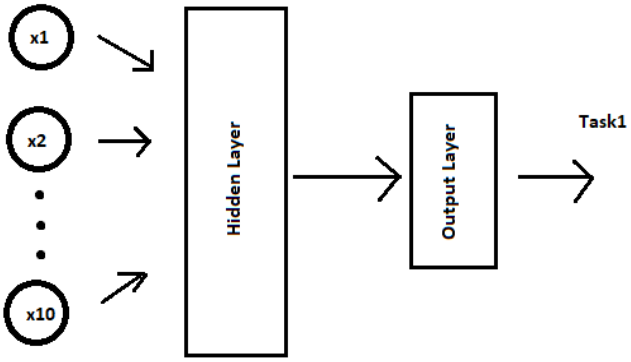

Fig 2. Single Task Learning

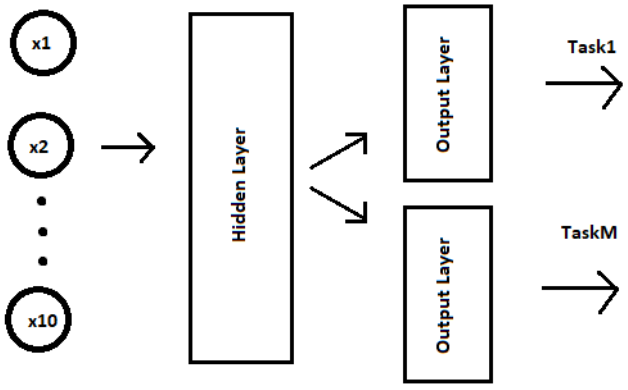

Fig 3. Multi Task Learning

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12324

The inputs are then again classified into various output categories. CNN architectures that would be used are: VGG, ResNet, Inception and MobileNet.

a. VGG-16 architecture

It is made up of many block of layers. It acts as a step by step process for performing the classification task. The various layers that are found in it are Convolution Layers, Max-Pooling Layer, 3 fully connected layers.

b. ResNet architecture

It overcomes problems of using residual network or learning blocks.

c. Inception - Google's CNN

This was developed to solve problems of object sizes. Inception has been updated to Inception-v2 and Inception-ResNet.

d. MobileNet architecture

This has been optimized for mobile and embedded applications with limited resources.

c. Proposed Convolution Neural Network

The proposed architecture is divided into two subsystems - feature extractor and modified classifier. Several CNN based architectures with similar brand and model classifications were used for investigation. Each baseline went through two stages. The results were compared from both the baseline & proposed architecture. The training was done and VGG-16 was not able to reach convergence. The second classification was done on aiming at model classification. It was more worse than previous classification. The modification done to improve the classification was done by replacing classifier part from the original VGG-16 to the proposed architecture. Then the make and model classification takes place via passing through two branches. Texture feature is used in max-pool layer [12]. There are dense layers and dropout layers. The activation function used is dense5 (SoftMax) To evaluate performance several methods were used, like accuracy scores, specificity, F-1 score. It provides detailed analysis of the performance. The other CNN architectures include: ResNet50 [13,14], ResNet101 [15], VGG-16 [16,17] and Inception [18,19]. The architecture of the proposed architecture is given in the Fig 4. below.

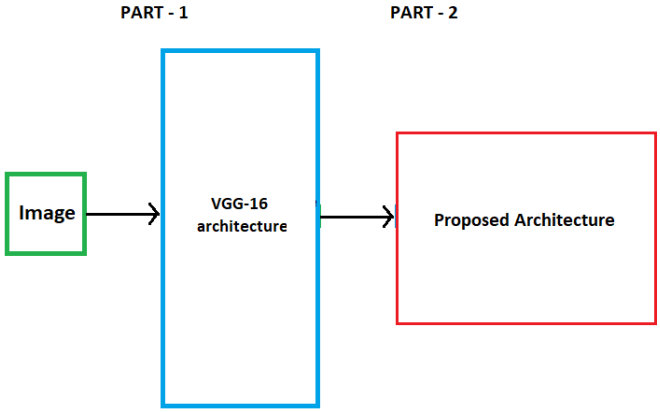

Fig 4. Proposed Architecture Outline

In the proposed architecture outline, there are two parts into which it is divided into, Part 1 and Part 2. In the Part 1 the image dataset is taken as input and passed to the VGG-16 architecture. In Part 2, the VGG-16 output is sent to the Proposed Architecture which implements the dense layer and dropout layer. These layers then do the classification task of both model and make classification in a separate outputs. Then the both outputs are taken into consideration and compared for the final analysis output. The detailed architecture of what the Part 1 and Part 2 consist of and the various parameters and how they are related is given in Fig 5. Part 1 VGG-16 Architecture and Fig 6. Part 2 Proposed Architecture below. The Part 1 VGG-16 Architecture consist of image input which is of size 224 x 224 x 3. This is the vehicle data input. It is passed to the VGG-16 architecture. There are a lot of layers in the VGG-16 architecture. There are 5 convolution layers which are further subdivided into different layers. In between the max pooling is also done. In convolution layer 1, there are two subdivisions. In both of these layers, the size of convolutional layer is (224, 224, 64). A max pooling (2,2) layer is passed after the convolution layer 1. In the convolution layer 2, there are two sub divisions just as the first convolutional layer. It has the layers dimensions as (112, 112, 128). Then a max pooling layer is passed of size (2,2). Third Convolutional Layer consists of 3 subdivisions which are of the dimensions (56, 56, 256). A max pooling layer (2,2) is added at the first subdivision of the convolution network layer 3. At the fourth convolutional network layer, the dimensions are (28, 28, 512), it also has 3 subdivisions. Max pooling layer is added after the third sub division. The last or 5th convolutional layer has the dimensions (14, 14, 512). It also has 3 subdivisions and it passes its output to the proposed architecture's first entity Feature Map. The Pooling is done in the original architecture of VGG-

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12324

16 but in this case, the proposed architecture applies some other layers to get a better result. This model is specifically designed for a better accurate classification where the vehicle types are very much similar. Thus, two branches are taken parallely into consideration for both vehicle model and vehicle make.

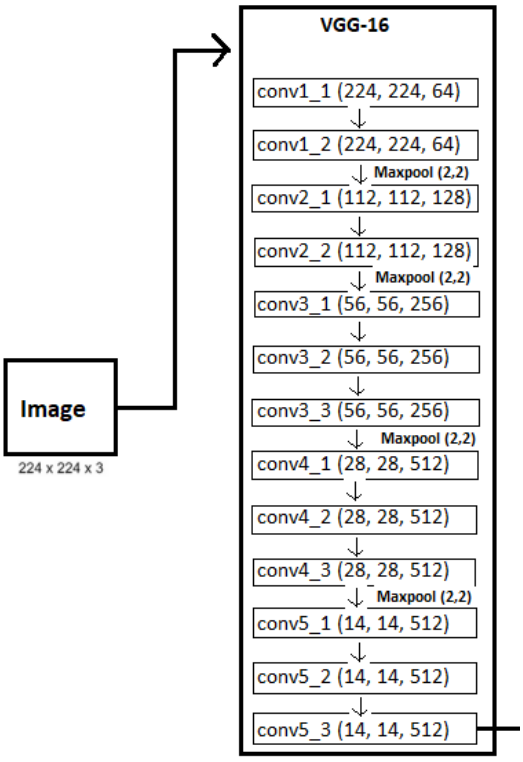

Fig 5. Part 1 VGG-16 Architecture

In the Part 2 of the architecture, the main aim is to modify the original architecture. It would make the improvement in the classification result. As show in Fig 6. The Part 2 of the architecture gets the input from the last layer of convolution network. It gets passed to the Feature Map of First step in the proposed architecture. The proposed architecture could be divided into two parts, the top layer and the bottom division layer. In the top layer, there is a feature map which takes the input from VGG-16 and then transfers it to Global Average Pooling (1, 512). There are two dense layers and two dropout layers. The dimensions of the dense layer is (1, 256) and that of dropout layer is 0, 1. After the second dropout layer, the model gets divided into two parallel branches. One for the model classification and another one for the make classification. The vehicle make is nothing but the brand classification. In each of the individual branches, there are 3 dense layers and 2 dropout layers. The dense layer consist of (1, 128) , (1, 64) intermediate layer dimensions. The difference happens at the last dense layer. For the vehicle model classification the layer dimensions are (1, 10) and the vehicle brand classification's final dense layer has the dimensions (1, 4). These steps classify both the model and brand in a very precise manner.

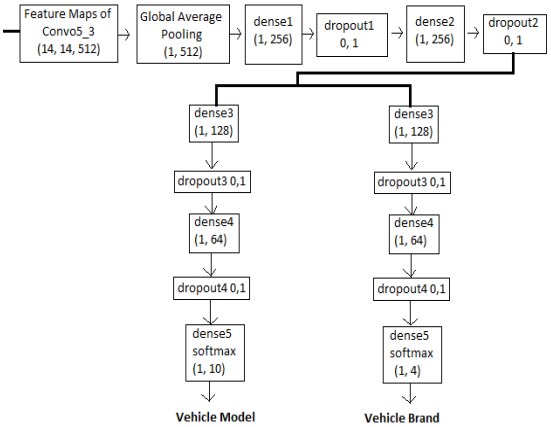

Fig 6. Proposed Architecture

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12324

IV. EVALUATING CLASSIFICATION MODELS

The following metrics are the best in evaluating classification models:-

1. Classification Accuracy: It is the most widely used criteria or classification evaluation. It is defined as the ratio of no. of correct predictions to that of overall no. of predictions. In terms of classification, MTL performs well by combining knowledge from several tasks [20]. But always accuracy is not the best metric via which the model is evaluated. It is because accuracy does not take into consideration imbalance datasets. One way to check for the accuracy as best metric is to construct a confusion matrix. It has True/False and Positive/Negative values. It is helpful in understanding precision and recall value. The representation of Confusion Matrix is given in Fig 7. The parameters that are present in it are

TN - True Negative
TP - True Positive
FN - False Negative
FP - False Positive

Confusion Matrix

		Predicted Value	
		Positive	Negative
Real Value	Positive	TP	FP
	Negative	FN	TN

Fig 7. Confusion Matrix

2. Precision: Precision defines the data that is truly classified as positive. The equation is given below in Fig 8.

$$Precision = (True\ Positive) / (True\ Positive + False\ Positive)$$

Fig 8. Precision Formula

3. Recall: It is defined as no. of true positives divided by the sum of true positives and false negatives. The Formula is given in Fig 9.

$$Recall = (True\ Positive) / (True\ Positive + False\ Negative)$$

Fig 9. Recall Formula

4. Specificity: .It is similar to sensitivity, only the focus is on the negative class. The formula is shown in Fig. 10

$$Specificity = (True\ Negative) / (True\ Negative + False\ Positive)$$

Fig 10. Specificity Formula

5. Fall-out: Fall-out determines the probability of determining a positive value when there is no positive value. The formula is shown in Fig 11.

$$Fall\text{-}out = (False\ Positive) / (True\ Negative + False\ Positive)$$

Fig 11. Fall-out Formula

6. Miss Rate: Miss Rate can be defined as proportion of positive values that were incorrectly classified as negative examples. The formula is shown in Fig 12.

$$Miss\ Rate = (False\ negative) / (True\ positive + False\ negative)$$

Fig 12. Miss Rate

7. Receiver Operator Curve (ROC curve): It displays the relationship between sensitivity and fall-out. The result of the performance is displayed in the form of curve.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023

DOI: 10.17148/IJARCCE.2023.12324

V. APPLICATIONS OF CLASSIFICATION

Some of the real life applications of classification are mentioned below along with the scope of it in that domain:

a) Email Spam: Email spam classification is done on the basis of the frequency and location of a set of spam words that most of the spam emails use. The model is trained on that using the vector data structure. The new emails are classified according to the model training and the output class would be either spam or non-spam.
b) Handwritten Digit Recognition: The classification is done for the input data which consist handwritten digit from 0-9. The model is trained on a large dataset of numbers and it classifies the new instance from among the set of range of possible output classes which is 0-9.
c) Image Segmentation: Classify the image into various segments based on the different entities present in it. This is done by calculating the area where similar type of data is found. This is then classified as a separate class output.
d) Speech Recognition: Speech recognition is a concept that identifies the voice data and then assigns it a class output, based on what the content that is said or conveyed in the voice input. After the recognition of the speech, then the corresponding actions could be performed.
e) DNA Expression Microarray: The disease or the tissue is identified based on the gene expression level. The cleaning process is happened on the input of DNA dataset and the output is the disease or tissue class output.
f) DNA Sequence Classification: The genomes are made up of DNA sequences, and every sequence (represented by A, C, G, T) is having a specific biological function. Classifications are the types of random process realizations.

CONCLUSION

The main aim of this paper is to give an insight to the reader regarding the basic fundamentals behind object detection and classification. It also mentions different types of object detection and classification algorithms. There are some applications of it for Intelligent Transport Systems (ITS), Smart Cities and Traffic Surveillience. It has some advantages/disadvantages which are mentioned in this paper. Classification using VGG-16 architecture have been used for the object detection and classification.

In addition to the basic overview regarding the object detection methods and classification algorithms by using the trained models after providing the dataset, this paper also gives the reader a brief overview of the challenges faced during the dataset collection as well as processing. The datasets might not be clear or faded, these have to be altered and modified such that they are usable and classifiable after the object gets detected. Some of the preprocessing methods are also used to make the vechicle illuminate if its night or highlight just the relevant fetures of the vehicle. The alteration of the images, flip horizontal and vertical and using the mirror image for training has also been taken into consideration.

These steps make the dataset ready for the usage and classifying more vehicles with further accuracy. The model is then tested with a new dataset and the classification accuracy score is noted. The evaluating classification models section sheds some light on this. Also, applications of classification is given at the end of the paper.

REFERENCES

[1] Akinyelu, A.A.; Zaccagna, F.; Grist, J.T.; Castelli, M.; Rundo, L. Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey. J. Imaging 2022, 8, 205.
[2] Ahmad, S.F.; Rahmat, M.K.; Mubarik, M.S.; Alam, M.M.; Hyder, S.I. Artificial Intelligence and Its Role in Education. Sustainability 2021, 13, 12902.
[3] Aligholi, S.; Khajavi, R.; Khandelwal, M.; Armaghani, D.J. Mineral Texture Identification Using Local Binary Patterns Equipped with a Classification and Recognition Updating System (CARUS). Sustainability 2022, 14, 11291.
[4] Abduljabbar, R.; Dia, H.; Liyanage, S.; Bagloee, S.A. Applications of Artificial Intelligence in Transport: An Overview. Sustainability 2019, 11, 189.
[5] Murali, A.; Nair, B.B.; Rao, S.N. Comparative Study of Different CNNs for Vehicle Classification. In Proceedings of the 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, India, 13–15 December 2018.
[6] Manzoor, M.A.; Morgan, Y. Vehicle Make and Model classification system using bag of SIFT features. In Proceedings of the 2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 9–11 January 2017.
[7] Abbas, A.F.; Sheikh, U.U.; Mohd, M.N.H. Recognition of vehicle make and model in low light conditions. Bull. Electr. Eng. Inform. 2020, 9, 550–557.

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁜ Impact Factor 7.918 ⁜ Vol. 12, Issue 3, March 2023
DOI: 10.17148/IJARCCE.2023.12324

[8] Manzoor, M.A.; Morgan, Y. Vehicle make and model recognition using random forest classification for intelligent transportation systems. In Proceedings of the 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA, 8–10 January 2018.

[9] K. Ying, A. Ameri, A. Truvedi, D. Ravindra, D. Patel, and M. Mozumdar, "Decision tree-based machine learning algorithm for innode vehicle classification," 2015 IEEE Green Energy and Systems Conference (IGESC), Long Beach, CA, 2015, pp.71-76. DOI:10.1109/IGESC.2015.7359454

[10] Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on pattern analysis and machine intelligence 24(7), 971–987 (2002)

[11] Fazli, Saeid, Shahram Mohammadi, and Morteza Rahmani. (2012) "Neural Network based Vehicle Classification for Intelligent Traffic Control", International Journal of Soft. Eng. & Applications Vol.3, No.3, pp. 17-22. DOI: 10.5121/ijsea.2012.3302.

[12] Chen, W.; Sun, Q.; Wang, J.; Dong, J.J.; Xu, C. A Novel Model Based on AdaBoost and Deep CNN for Vehicle Classification. IEEE Access 2018, 6, 60445–60455.

[13] Boukerche, A.; Ma, X. A Novel Smart Lightweight Visual Attention Model for Fine-Grained Vehicle Recognition. IEEE Trans. Intell. Transp. Syst. 2021, 28, 1–17.

[14] Siddiqui, A.J.; Boukerche, A. A Novel Lightweight Defense Method Against Adversarial Patches-Based Attacks on Automated Vehicle Make and Model Recognition Systems. J. Netw. Syst. Manag. 2021, 29, 41.

[15] Ma, X.; Boukerche, A. An AI-based Visual Attention Model for Vehicle Make and Model Recognition. In Proceedings of the 2020 IEEE Symposium on Computers and Communications (ISCC), Rennes, France, 7–10 July 2020.

[16] Naseer, S.; Shah, S.M.A.; Aziz, S.; Khan, M.U.; Iqtidar, K. Vehicle Make and Model Recognition using Deep Transfer Learning and Support Vector Machines. In Proceedings of the 2020 IEEE 23rd International Multitopic Conference (INMIC), Bahawalpur, Pakistan, 5–7 November 2020.

[17] Liu, D.; Wang, Y. Monza: Image Classification of Vehicle Make and Model Using Convolutional Neural Networks and Transfer Learning; Stanford University: Stanford, CA, USA, 2017

[18] Balci, B.; Elihos, A.; Turan, M.; Alkan, B.; Artan, Y. Front-View Vehicle Make and Model Recognition on Night-Time NIR Camera Images. In Proceedings of the 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Taipei, Taiwan, 18–21 September 2019.

[19] Balci, B.; Artan, Y. Few-Shot Learning for Vehicle Make & Model Recognition: Weight Imprinting vs. Nearest Class Mean Classifiers. In Proceedings of the 2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC), Rhodes, Greece, 20–23 September 2020.

[20] Zhang, Y.; Yang, Q. An overview of multi-task learning. Natl. Sci. Rev. 2017, 5, 30–43.