Sorokh-Poth: A Balanced Bangladeshi Road Vehicle Image Dataset Integrated With a Detection System Based On Deep Convolutional Neural Networks

Tamanna Zubairi Sana[a], Sabiul Islam[b], Nazmul Hasan[c], Istiak Ahmad[d]

[a] tamannasana44@gmail.com
[a] Department of Computer Science and Engineering, East Delta University, Chattogram 4209, Bangladesh
[b] Department of Computer Science and Engineering, United International University, Dhaka 1212, Bangladesh
[c] Department of Computer Science and Engineering, Daffodil International University, Dhaka 1216, Bangladesh
[d] Department of Computer Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Abstract

When it comes to the administration of highways, the identification of intelligent vehicles is becoming an increasingly important component. Applications for vehicle detection are essential for both military and civilian usage, including the monitoring of highway traffic, administration, and the control of urban traffic. The vast amount of vehicle image data available online has the potential to encourage the development of increasingly sophisticated object recognition and classification models and algorithms. But finding an organized, balanced, and meaningful dataset continues to be a major challenge. This study proposes "Sorokh-Poth," a new complete balanced image dataset based on Bangladeshi road travel that is compatible with a number of CNN-based architectures. The majority of the photographs in the collection were taken with a smartphone. The dataset includes 9,809 classified and annotated images of ten different types of vehicles, including autorickshaws, bicycles, buses, cars, CNG-powered vehicles, lagoon rickshaws, trucks, and vans. This research work utilizes the residual network ResNet-50 model, a CNN-based architecture. Here, features specific to the type of vehicle were automatically retrieved and grouped. Accuracy, precision, recall, and f1-score were just a few of the metrics used for evaluation during the research. The proposed model exhibited an increasing accuracy despite the vehicles' shifting physical characteristics. The purpose of this work is to implement various CNN-based residual networks ResNet-50 and ResNet-150V2. Our proposed ResNet-50 model achieves an accuracy of 98.00% in the detection of native Bangladeshi vehicles, according to result comparisons and evaluations.

Keywords: Sorokh-Poth; ResNet-50; deep learning; ResNet-150V2; transfer learning; Bangladeshi vehicle classification

1. Introduction

The most popular route taken for short, daily activity distances is the road. Due to the rising number of fatalities and severe injuries caused by road traffic accidents, this issue is one that affects the entire world. The developing world accounts for more than 80% of all traffic injuries. When engine-driven cars began to dominate the global transportation network on the highways, the necessity for traffic management and navigation systems first became apparent. In order to address this demand, several traffic rules were created,

and various traffic signs were developed for use on the sides of the roads. These measures were designed to guide drivers and promote safe driving in order to reduce the amount of traffic accidents, backups, and other problems. Some first-world countries currently use a variety of smart car technologies to supplement manual human navigation as traffic navigation technology has evolved. Consequently, automated navigation systems like traffic sign detection and interpretation as well as voice command processing methods are employed [1]. In a prospering nation like Bangladesh, the amount of vehicles that are driven on the roads is growing at an alarming rate, but the traffic control system has not been upgraded.

One of the few developing countries with a very high rate of accidents, injuries, and fatalities is Bangladesh. The current state of affairs is the worst ever. As of October 2020, the country had more over 4.4 million registered automobiles, according to the Bangladesh Road Transport Authority (BRTA) [2]. Because the majority of the drivers and assistants on local transportation are so ignorant or unaware, they cannot or will not obey traffic signs. Due to this, major cities like Dhaka, Chittagong, etc. now frequently experience traffic violations, violent accidents, and intolerable traffic congestion. It has evolved into one of the most significant issues we deal with on a daily basis. Therefore, the necessity for an autonomous navigation aid on the vehicles that can detect and understand traffic signs on the sides of the road and help drivers navigate safely is even larger than it is in most other nations across the world.

According to statistics, almost 1.2 million individuals pass away on the roads each year. Additionally, estimates indicate that approximately 50 million individuals are hurt in traffic accidents worldwide [3]. However, the occurrence varies by country. In underdeveloped nations, injuries, deaths, including tragedies are ten to sixty times more prevalent than in developed countries. In order to meet the Sustainable Development Goals (SDGs) (2030), Bangladesh is expected to drastically reduce the amount of people who are killed or injured as a result of congested roads incidents. Notwithstanding this, Bangladesh must strengthen its traffic control system as well as rely on a smart transportation network to stay up with wealthy nations.

Vehicle detection is a significant part of automated driving and traffic surveillance. Traditional vehicle detection methods have shown promise, such as the Gaussian mixed model (GMM) [4]. But because of the varying lighting, background clutter, occlusion, etc., it is not perfect. Vehicle detection continues to be a significant computer vision task.

The vast amount of vehicle picture data available online has the potential to encourage the development of increasingly sophisticated object recognition and classification models and algorithms. Organized, balanced, and valuable datasets, however, continue to be a major issue for Bangladesh. Our primary goal in this study is to propose a brand-new, comprehensive, balanced image dataset called "Sorokh-Poth" [5] that is based on Bangladesh's road transportation system. We have taken pictures of live traffic signs using a camera. Our detection method has been thoroughly tested using a dataset that includes pictures we've taken under various lighting and environmental situations.

1.1. Aims and Objectives

The datasets used to categorize vehicles currently in use, such as Stanford Cars and CompCars, are somewhat tiny. These are based on specific geographic categorization. Thus, to obtain reliable final results, a competent dataset technique and vehicle detection in Bangladesh are needed. The study's goals are to address the following issues with Bangladeshi native vehicle classification systems:

- To improve upon the present datasets class, feature, and form limitations by introducing the "Sorokh-Poth," a native Bangladeshi vehicle dataset.
- To strengthen the adaptability of local Bangladeshi vehicle classification methods in both good and poor lighting conditions.

- To propose a novel, efficient approach for identifying and detecting vehicles of Bangladesh using deep learning techniques.
- To compare the outcomes of the various utilized vehicle classification models by using performance measures such as accuracy, recall, precision and f1-score.

The remaining sections of this work are organized as follows: In section 2, the previous studies that are relevant to the study that has been presented. Section 3, dive into the overall architecture of the model that has been suggested for the identification and classification of automobiles using the deep learning algorithm. In section 4, the experimental design and its analysis are presented. Section 5 addresses the result and output of the research. The conclusion and future works of the research are discussed in section 6.

2. Literature Review

Numerous approaches are already in use for the detection and identification of vehicles. The majority of these techniques have a variety of drawbacks, making them less than ideal for use with intelligent automation technology. This section includes a discussion of several relevant works on various techniques for vehicle detection.

Chen and Li [1] employed the deep learning method to investigate the vehicle identification algorithm, which employs the YOLOv3 algorithm as well as SSD algorithm as main targeted detection approaches. The image data from the accessible road vehicle dataset is initially processed by the technique as training data. The YOLOv3 and SSD algorithms are used in the construction of the vehicle detection model so that the impact of detecting may be demonstrated. The result is determined by contrasting the effect of the two approaches on vehicle detection. A conclusion was drawn based on the findings of the evaluation and emphasized the qualities of multiple models developed during training that are relevant to semantic segmentation, target monitoring, including automated vehicles.

Khalifa et al. [2] used CNN for the identification of autos from roadside webcam outputs to apply video processing techniques as well as retrieve the essential information. The article utilizes the YOLOv5s framework in conjunction with the k-means approach to optimize anchor boxes under varied illumination conditions. In accordance with the findings of the assessed algorithm, the suggested framework was capable of achieving a mAP of 95.1 in the night time and 97.8 in the daylight datasets.

Humayun et al. [3] in their study, evaluated the identification of automobiles in a scene under various weather circumstances, such as fog, sand and dust storms, wintry and wet weather both throughout the day as well as night. Their proposed design includes a layer for spatial pyramid pooling (SPP-NET) plus minimal layers for batch normalization to the CSPDarknet53 basic architecture. They also improved the DAWN Dataset with Saturation, Hue, Brightness, Exposure, Blur, Darkness, as well as Noise changes. This significantly expands their dataset, and also makes identification exceedingly tough. Throughout training, the model achieved an average accuracy of 81% and accurately identified the smallest car in the images.

Ozturk and Cavas [4] proposed a CNN based hybrid vehicle detection approach with the help from morphological operations. The results of their study on the COWC dataset demonstrated that, despite having just over 100,000 neural network parameters, their suggested architecture can recognize automobiles with a precision of 96%. Their proposed design uses a much smaller number of parameters as compared to commonly used neural networks, according to a comparison of parameter counts. The evaluation of performance and comparison of their suggested approach with YOLO-v4 demonstrated the efficiency of the suggested tiny CNN network.

To increase the robustness of vehicle categorization in real-time applications, Butt et al. [6] suggested a convolutional neural network-based approach. They offered a vehicle dataset that was made up of 10,000 images and was separated into the six most prevalent categories of vehicles. This was done in order to increase the reliability of real-time vehicle classification systems. Their research begins by doing initial fine-tuning on a self-built vehicle dataset using pretrained versions of AlexNet, VGG, Inception-v3, ResNet, and GoogleNet in order to evaluate the usefulness of these networks in terms of convergence and accuracy. Because of this, the suggested method acquired a precision level of 99.68%, an accuracy level of 99.65%, and an F1-score of 99.56% when utilizing the dataset that they had constructed themselves.

Sindhu et al. [7] proposed a study which works with the notion of Vehicle Detection with the support of Computer Vision algorithm in real-time frame utilizing continual video stream from CCTV from all around the universe. Their suggested framework uses YOLOv4 to accelerate real-time object recognition, and it was tested under a variety of situations, including low visibility, daylight, rain, snow and night.

In the headlight control system, a deep learning-based image recognition is suggested in the research proposed by Huang et al. [8]. Vehicles equipped with this system can judge at night, compute the safety distance between the vehicles in front, and detect the vehicle in front in real time while the driver is operating the vehicle.

The classification as well as recognition of cars traveling on highway networks have a substantial bearing on the management of traffic and the prevention of accidents. Chauhan et al. [9] came up with the concept of CNN framework, which is centers on the automobile classification algorithm at its core. This architecture was conceived with the intention of numbering vehicles as well as classifying them along significant routes. Their proposed model attains 75% MAP after adding 5562 videos to roadways.

In contrast to the region-based convolutional network method, Ma et al. [10] in their research offered a novel rotation-invariant vehicle recognition method that is precise, stable, and has a straightforward structure. The Munich vehicle dataset and the UAVDT dataset were used as the basis for their study. The experiment's findings show that the suggested approach performs satisfactorily.

Wang et al. [11] created an efficient R-CNN framework with primary emphasis on vehicle classification techniques. The objective was to provide a method for real-time traffic surveillance. 60,000 photos were included in the sample that the authors tested. This information was acquired and divided into trainable and testable sets. The total percent of accurate answers was 80.051%.

A GoogLeNet infrastructure for vehicle detection based on transfer learning in road traffic has been presented by Jo et al. [12]. The authors' experiments on the ILSVRC-2012 dataset revealed that the provided classifier has an accuracy rate of 0.983.

The research carried out by Kim et al. [13] made use of the PCANeT-HOG-HU model, the primary emphasis of which was the procedure of acquiring collective features. In this particular instance, the technique was utilized as data input for the SVM, which was used to train the classifier design. The researchers used surveillance footage to take 13,700 images of vehicles to use for both the training and testing of the proposed classifier model. There were six distinct types of automobiles included in this sample. Their suggested architecture for the detection of light to heavy loads has an accuracy level that averages out to 98.34%.

3. Methodology

Intelligent vehicle recognition is becoming incredibly valuable in highway regulation. Moreover, because vehicles come in a variety of sizes, it might be difficult to detect them. Applications for vehicle recognition is crucial for both civilian and military uses, including highway traffic surveillance, management, and urban traffic planning. The method of detecting vehicles on the road can be utilized for a variety of purposes, including vehicle tracking, counting the number of vehicles on the road, determining the average speed of each vehicle, conducting traffic analysis, and classifying the types of vehicles on the road. This study will concentrate on the type of vehicle classification on roads to address the shortcomings so that various

countries, particularly South Asian countries like Bangladesh, might benefit from its adoption. These nations continue to employ traditional methods that are manually monitored by programs that use sensors, photos, and people. To obtain reliable findings, the Bangladeshi traffic surveillance system must be operated competently. Our principal targets are to provide a well-established dataset of Bangladeshi vehicles and to develop a system that can correctly classify the vehicles.

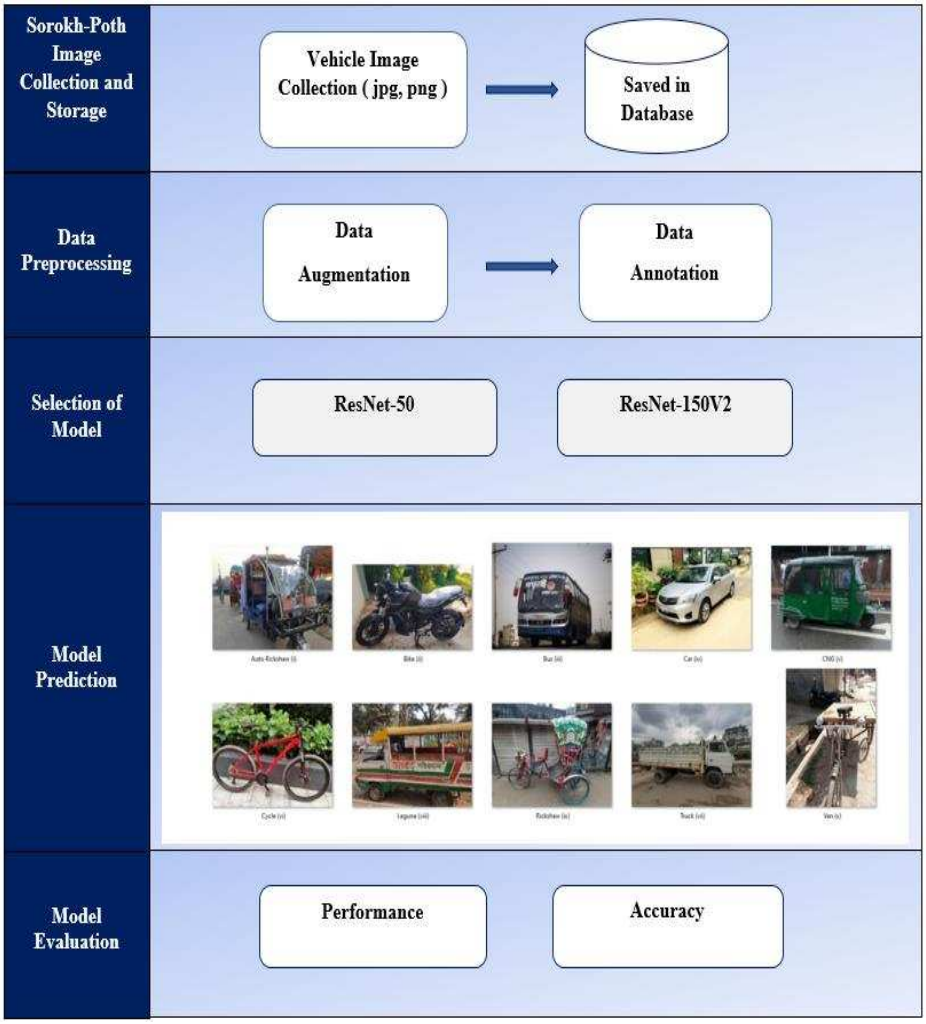

Fig. 1. Proposed System Architecture of Bangladeshi Vehicle Detection

In this research, we put forward an approach for the classification as well as the detection of Bangladeshi native vehicle types that is rooted on transfer learning, integrating data augmentation. With CNN's tremendous advancement in image processing, there are several different CNN designs that excel at image classification. We selected ResNet-50 and ResNet-150V2 as candidate networks due to their outstanding

classification performance in ImageNet. With the aid of transfer learning and data augmentation approaches, we suggested a DL model for locating native Bangladeshi vehicles.

3.1. Sorokh-Poth Dataset

The Sorokh-Poth dataset is a comprehensive collection of road vehicle image dataset including 10 distinct categories like Auto-rickshaw, bike, bus, bicycle, car, CNG, leguna, rickshaw, truck, van. This dataset has a total number of 19,636 images and annotated files. The proposed data source of Sorokh-Poth dataset also differs from the ones that are readily available. There is a lot of variety in terms of traits and composition. Furthermore, because it was created using very small data samples with few classifications, the current vehicle classifier might not function properly in actual ITS deployments.

The dataset contains 4 folders and each folder contain a total of 10 folders for vehicles categories as follows:

Fig. 2. Sample images of different categories

- **Raw Image:** The dataset contains a total of 4951 raw images for 10 distinct categories.
- **Raw Image Annotation:** The dataset contains a total of 4951 annotation files for the corresponding raw image.
- **Data Augmented Image:** There are 4867 images are generated from the raw image by using several data augmentation techniques.
- **Data Augmented Image Annotation:** This folder contains 4867 annotations files for corresponding augmented images. The annotation files save in XML format.

3.2. Dataset Collection

In Bangladesh, traffic management is one of the central quandaries which has an adverse influence on national and social development. To mitigate the concern, we require to build an automatic Artificial intelligence-based system that can assist us to dominate the traffic system properly. To promote an AI-based system, we require a considerable amount of data that can be used to train the system properly. The Sorokh-Poth dataset is specifically built for the road management system. We assembled the image using the smart phone and social media. During capturing the images, we consider the different weather condition, background, direction of the vehicle, and light condition etc. We collected a total of 5000 images by smartphone and social media like Facebook.

Table 1. Distribution of Image and XML files

Categories	Raw Image Files	Raw Image Annotation Files	Augmented Image Files	Augmented Image Annotation Files	Total
Autorickshaw	502	502	502	502	2008
Bike	502	502	505	502	2014
Bus	591	591	590	590	2362
Car	568	568	554	554	2244
CNG	524	524	517	517	2082
Leguna	321	321	320	320	1282
Rickshaw	375	375	366	366	1482
Truck	537	537	537	537	2148
VAN	519	519	464	464	1966
Total	**4951**	**4951**	**4867**	**4867**	**19636**

3.3. Data Preprocessing

Data preprocessing is very necessary in order to construct a classifier based on deep learning, such as one for the detection of vehicles. This is because the images of the vehicles are obtained from a variety of sources. Following this, data preprocessing is carried out to reduce the amount of noise or undesirable background, shrink the standard format image, and construct the images of the vehicles with an uneven lighting system. In the proposed research, the preprocessing step is comprised of two distinct phases.

3.3.1. Data Augmentation

Image data augmentation approaches have the potential to produce mutations of the images, which can improve the capacity of deep learning models to apply what they have learned to a variety of diverse images. Image augmentation may be accomplished in a variety of ways, including by moving the image horizontally and vertically, flipping it horizontally and vertically, rotating it, adjusting the brightness, cropping it, scaling it, zooming it, and adding Gaussian noise to it. In order to effectively balance the dataset, many methods like shifting, flipping, rotating, and zooming are employed, and 5000 augmented photos are created.

3.3.2. Data Annotation

Image annotation is a sort of data labeling often referred as transcribing, tagging, or processing. It also called metadata of the data. To annotate the images, the popular annotation tool Labeling by Tzuta Lin is used. The annotation process is divided into five steps: 1) Open the image by using Labeling tools. 2) Manually draw a rectangle frame to the boundary of the target object to define its exact position in that image by X-Y coordinates. 3) Assign the class of the object. 4) Finally, save the XML file in PASCAL VOC format. A total of 10000 images is annotated which contain both raw images and augmented images distinctly.

3.4. Convolutional Neural Network (CNN) Model

CNN is indeed a deep learning model that really can comprehend data such as photographs that have a grid pattern. CNN was created with the animal visual brain in mind, which is designed to gain spatial hierarchies of features by traversing lower to upper-level components in an efficient and dynamic manner. The goal of CNN is to achieve this. Convolutional, pooling and fully connected layers are the three sorts of variations of layer types that are often seen in a traditional CNN. Convolution and pooling layers are applied in the process of feature extraction, and a fully connected layer turns the collected features into outputs that include categorization. The convolution layer, which is composed of a series of arithmetic calculations, a specific type of linear operation, is critical. Since a feature can be located anywhere in a digitized picture, the pixel values are saved in an array of integers or a two-dimensional grid. Moreover, a relatively small grid of parameters known as the kernel, which is an optimizable feature extractor, is adjusted at each image point. As a result, CNNs are particularly effective for image processing. CNN for its superior image detection capabilities is chosen since it is needed to properly identify the vehicles.

- Convolution Layer

It's the first layer, and its purpose is to isolate the distinguishing characteristics of the input visuals. In this layer, the scientific process known as convolution takes place. It involves the source image as well as a filter with a particular size, denoted by the notation MxM. The dot product is calculated between the filters and the components of the source image while the filters are moved smoothly across the source image. The size of the filter has an effect on how accurately the dot product is calculated (MxM). The Feature map that was produced as a consequence stores information regarding the image, including its borders, edges, as well as corners. After that, this feature map is put to use to instruct future layers on additional characteristics taken from the original image. Upon the completion of a convolution operation on the input, the output is passed on to the subsequent layer by the convolution layer. Because of CNN's convolutional layers, the spatial connection between both pixels is preserved.

- Pooling Layer

The major objective of the pooling layer is to lessen the magnitude of the convolved feature map so as to cut down on the amount of complexity caused by computing. This is accomplished individually on each feature map and so by decreasing layer linkages. Based on the method used, there are various types of pooling procedures. The feature map yields the greatest component in Max Pooling. The average pooling algorithm determines mean value of the image's components within a specific segment size. The Sum Pooling operation calculates the total value of the items contained in the particular segment. The Convolutional Layer and the FC are frequently connected through the Pooling Layer. This CNN model generalizes the features that were collected from the convolution layer, which enables the networks to recognize the characteristics on their own. This helps to reduce the number of calculations that take place within a network.

• Fully Connected Network

To link neurons that are located on different layers, the Fully Connected (FC) layer is utilized. This layer also takes into account biases and weights. These are the layers that come immediately before the output layer in a CNN's architecture because they are the penultimate few layers before the output layer. This layer is responsible for the compression and transmission of the input image that was received from the layers below to the FC layer. After then, the compressed vector would be sent to a few more FC levels, which are essentially just places where standard mathematical computation operations are carried out. The process of classifying starts at this moment. The overall performance of two entirely linked layers is greatly increased over that of a single connected layer, which is why two layers are connected and joined together. As a direct result of these CNN layers, the amount of human supervision that is required has been reduced.

• Dropout

Overfitting occurs in the training sample when all features are connected to the FC layer. It appears when a certain model works exceptionally well on the training data that it results in a negative impact on the model's efficiency when deployed to fresh data. Overfitting may also occur when a model performs very well on test data. A dropout layer is developed to address this issue. When the neural network is trained, this layer removes a tiny fraction of its neurons, lowering the model's complexity. After completing a dropout of 0.3, an arbitrary 30% of the nodes inside the neural network are deleted. Neurons are eliminated from neural networks throughout the training to create place for new ones.

• Activation Functions

The CNN paradigm relies heavily on the activation function as a fundamental building block. They are utilized to locate and approximately determine any continuous as well as intricate relationship amongst the variables of a network. In layman's words, it specifies which modeling info should and should not be transmitted across the network. As a consequence, the structure becomes nonlinear. Some of the most common activation functions are the Softmax, tanH, ReLU, and Sigmoid functions. Every one of these activities serves a specific purpose. For binary classification, the sigmoid as well as softmax functions are preferred, whereas softmax is often used for multi-class categorization.

3.5. ResNet

ResNet which stands for Residual Network well-known deep learning model. Over the years, deep convolutional neural networks have achieved major advancements in image recognition and categorization. Digging further has found effective for handling with progressively tough circumstances including improving categorization or detection rates. Furthermore, concerns such as the deterioration issue and also the diminishing gradient issue has made training deeper neural networks difficult. Both of these concerns are addressed by residual learning. Every layer in a neural network is trained for the current task whilst acquiring low- as well as high-level information. With residual learning, a model attempts to learn certain residual instead of all features. The activation is finished whereas the input 'x' is added as just a remnant to the weighted layers' output. The model makes use of relu activations. ResNet-50 is a residual network with 50 layers that also comes in ResNet-101 and ResNet-150 variations. The ResNet-50 and ResNet-150V2 networks are utilized in this research for the classification of vehicle images. Using both of these networks as a pre-trained model, successful outcomes for the classification task is obtained.

3.5.1. ResNet-50

The deep neural networks' performance stagnates or worsens as more layers to them is added. The cause of this is the vanishing gradient issue. Gradients become vanishingly small as a result of back propagation through the deep neural network and repeated multiplication, which results in the issue. ResNet can solve the vanishing gradient problem by using Identity shortcut connections and frequently skip connections which bypass one or even more levels. Layer N+Z input and layer N output are connected by shortcut connections. ResNet-50 is made up of 48 convoluted layers, one maximum pooling layer, as well as an average pooling layer. There seem to be 50 weighted layers, as well as 25,583 592 trainable features. For the purpose of image classification, the proposed system primarily relied on a ResNet-50 neural network, which had been pre-trained mostly using the ImageNet database. In order to relocate the first 49 layers of the ResNet-50 model, which had been blocked on the classification algorithm, transfer learning approaches were applied.

3.5.2. ResNet-150V2

One of the most potent convolution neural networks, residual nets use skip connections to solve the vanishing/exploding gradients issue. Bypassing the intermediate levels, the skip connection allows activations on one layer to be connected to those on succeeding layers. As a result, a leftover block is created. These residual blocks must be stacked one after the other in order for ResNet's to form. In order to reduce training time and model complexity at greater depths, the 150V2 design uses a stack of 6 layers. To improve accuracy and reduce validation error, two densely linked layers is added with drop-out on top of the original model.

3.6. Model Selection and Training

A deep learning technique forecasts a model based on the output "Y" as well as an attribute map to the target "X" as input data. The ResNet-50 framework was employed for our model. The factors (update biases and weights) that were employed for our model's recognition were modified by the algorithm while training. In the proposed research, 80% image data is utilized for training the model. Furthermore, 20% of the image data was set aside to generate a validation subset to evaluate the model. The effectiveness of the deep learning model that was proposed was evaluated with the help of the test dataset.

4. Experimental Analysis and Results

In this research, data is splitted into two groups: the training set, which consisted of 80% of the image data, and the testing set, which consisted of the remaining 20% of the image data. For the classification of vehicles, we used the ResNet-50 and ResNet-150V2 networks. The most accurate model's assessment metrics, such as Accuracy, Classification Report, Confusion matrix, Precision, Recall, and so on, were used to approve the expected performance of the model for the training data.

4.1. Confusion Matrix

A confusion matrix is a table that is used to assess the performance of a classification system. A confusion matrix depicts and summarizes a classification algorithm's performance. The confusion matrix is comprised of four primary attributes, each represented by a number, which are utilized to supply the classifier with its parameters of measurement. The four properties are: TP (True Positive), False Positive (FP), True Negative (TN) and False Negative (FN). By creating assessment criteria based on these four primary properties, the effectiveness of the suggested native recognition and classification approach is assessed.

The previously stated TP, TN, FP, and FN are the building blocks upon which the algorithm performance measurements known as accuracy, precision, recall, as well as F1 score are constructed. The instances in each row of the matrix correspond to the actual class, whereas the instances in each column correspond to the predicted class. In this matrix of prediction outcomes, the identification accuracy is represented by a blue colorbox, and also the darker the hue, the further successful the model detection. The expected values of the test set, including autorickshaw, bicycle, bike, bus, car, CNG, leguna, rickshaw, truck, and van are represented on the horizontal axis. The authentic values of the test sets, including the autorickshaw, bicycle, bike, bus, car, CNG, leguna, rickshaw, truck, and van are represented by the vertical axis. On the crosswise axis of the matrix is the model's estimated value that agrees with the actual value of the test set. The results of the confusion matrix of the ResNet-50 model is shown in Figure 3.

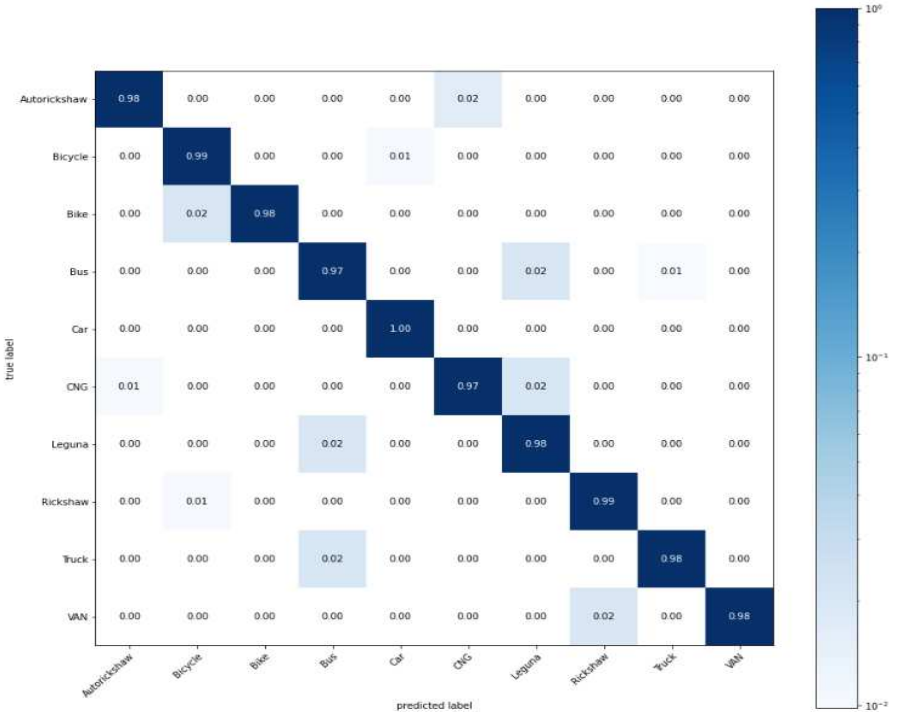

Fig. 3. Confusion Matrix of ResNet-50

4.2. Classification Report

An evaluation of the performance of machine learning is presented in this study. It is used to demonstrate the Precision, Recall, F1 Score, and Support of the trained classification model.

4.2.1. Precision

Precision is the events that were designed to maximize the degree of precise and definitive agreements (true positive). It now offers not only the total number of true positive a false positive traits, but also the proportion of characteristics that are true positives.

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive} \quad (1)$$

4.2.2. Recall

The recall of a classifier is its ability to detect all positive events. For each class, it is described as the proportion of true positive instances to the total of true positives as well as false negatives instances.

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative} \quad (2)$$

4.2.3 F1 Score

The congruent mean of accuracy and recall is used to get the F1 score. Therefore, accuracy and recall are combined into a single calculation. It can be said in the following way:

$$\text{F1 Score} = 2 * \frac{(Recall * Precision)}{(Recall + Precision)} \quad (3)$$

The classification report of the proposed system model is illustrated below:

Fig. 4. Classification Report of ResNet-50

5. Results and Discussion

The ResNet-50 and ResNet-150V2 models were loaded using TensorFlow resources for evaluation. All networks were trained using TensorFlow. The Adam optimizer, a stochastic optimization method, is used to optimize the parameters in our suggested native vehicle classification model. For both dropout layers, dropout ratios of 0.50 is used, and learning rate at 0.0001 is settled. 50 epochs and a batch size of 32 were used to train the model. A commonly used loss function, categorical cross-entropy, was employed to accumulate loss throughout the process, and validation of the network was carried out after each epoch to assess the learning.

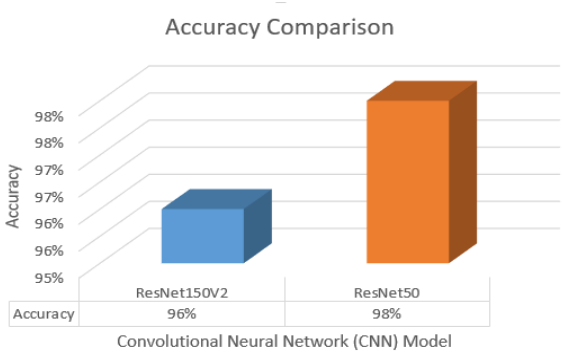

Fig. 5. Accuracy Comparison of CNN Networks

As shown in figure 4.3, the proposed approach achieved validation accuracy of 96% for ResNet-150V2 and 98% for the ResNet-50 pre-trained model despite having been trained for 50 epochs.

Table 2. Performance Comparison Report

Model	Precision	Recall	F1 Score	Accuracy
ResNet-150V2	0.942	0.963	0.963	0.96
ResNet-50	0.978	0.982	0.982	0.98

The performance comparison report is given in table 2. It is clearly observed that the ResNet-50 model performed more accurately than ResNet-150V2, with an average difference of 2%. As a result, after the architecture is tweaked, ResNet-50 is anticipated to achieve more accuracy. Since it is clearly evident that the suggested ResNet-50 pre-trained deep learning model, which is rooted on transfer learning, outperforms the other classification approach ResNet-150V2 in regards to accuracy in classification, it is possible to draw the conclusion that the proposed model, which is based on ResNet-50, can extract robustness for recognition as well as classification in case of Bangladeshi native vehicles that it is more reliable.

6. Conclusion & Future Works

The research that has been proposed is notable in a wide variety of different ways. As compared to other developing countries, the percentage of people killed or injured in road accidents in Bangladesh is substantially greater than in those other countries. For the purpose of resolving this matter, the Sorokh-Poth dataset was established. This dataset consists of 19,636 photos that are divided into 10 categories for the purpose of training the classification system. The training classification technique for native Bangladeshi cars uses this dataset. Initially, our Sorokh-Poth vehicle dataset was put through rigors training on two advanced CNN networks known as ResNet-50 and ResNet-150V2. These networks were trained to validate the performance of the dataset. Transfer learning was utilized in the development of the ResNet-50 model, which was then used for the categorization of automobiles that are indigenous to Bangladesh. The ResNet-50 architecture was improved via transfer learning by adding a new classification layer on top of the initial network. This layer was placed on top of the architecture. The performance of the model was assessed using a variety of criteria, including accuracy, precision, recall, and F1 Score, among others. The outcomes of the study demonstrated that the native vehicle categorization approach presented by the study had an accuracy of 98% when utilizing ResNet-50.

6.1. Future Scopes

The project's future prospective scopes focus on obtaining more exact findings by aiming to upgrade the existing dataset and include additional categories, such as people and traffic signs. This will allow for the project to reach its full potential in the future. It is also planned to make use of premium cloud training capabilities so that we may get infinite training time and further improve our outcomes. As was said before, the study that we are conducting is an offshoot of previous work, but it is extremely important both for the development of autonomous cars and for the regulation and monitoring of traffic. It is also planned that in the future, the system will be extended to include modules for automatic license plate recognition, traffic congestion detection, and vehicle counting. This will be done prior to combining these individual components into a comprehensive autonomous traffic monitoring system. One other possibility could be to use the suggested findings to the development of intelligent parking systems.

References

[1] Y. Chen and Z. Li, “An effective approach of vehicle detection using Deep Learning,” Computational Intelligence and Neuroscience, vol. 2022, pp. 1–9, 2022.
[2] O. O. Khalifa, M. H. Wajdi, R. A. Saeed, A. H. Hashim, M. Z. Ahmed, and E. S. Ali, “Vehicle detection for vision-based intelligent transportation systems using convolutional neural network algorithm,” Journal of Advanced Transportation, vol. 2022, pp. 1–11, 2022.
[3] M. Humayun, F. Ashfaq, N. Z. Jhanjhi, and M. K. Alsadun, “Traffic management: Multi-scale vehicle detection in varying weather conditions using Yolov4 and Spatial Pyramid Pooling Network,” Electronics, vol. 11, no. 17, p. 2748, 2022.
[4] M. Ozturk and E. Cavus, “Vehicle detection in aerial imaginary using a miniature CNN architecture,” 2021 International Conference on Innovations in Intelligent Systems and Applications (INISTA), 2021.
[5] Hasan, Nazmul ; Sana, Tamanna Zubairi; Islam, Sabiul; Ahmad, Istiak (2023), “Sorokh-Poth: A Balanced Bangladeshi Road Vehicle Image Dataset”, Mendeley Data, V1, doi: 10.17632/7xvcvxgphb.1
[6] M. A. Butt, A. M. Khattak, S. Shafique, B. Hayat, S. Abid, K.-I. Kim, M. W. Ayub, A. Sajid, and A. Adnan, “Convolutional neural network based vehicle classification in adverse illuminous conditions for Intelligent Transportation Systems,” Complexity, vol. 2021, pp. 1–11, 2021.
[7] V. S. Sindhu, “Vehicle identification from traffic video surveillance using Yolov4,” 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021.

[8] Z.-H. Huang, C.-M. Wang, W.-C. Wu, and W.-S. Jhang, "Application of vehicle detection based on deep learning in headlight control," 2020 International Symposium on Computer, Consumer and Control (IS3C), 2020.
[9] M. S. Chauhan, A. Singh, M. Khemka, A. Prateek, and R. Sen, "Embedded CNN based vehicle classification and counting in non-laned road traffic," Proceedings of the Tenth International Conference on Information and Communication Technologies and Development, 2019.
[10] B. Ma, Z. Liu, F. Jiang, Y. Yan, J. Yuan, and S. Bu, "Vehicle detection in aerial images using rotation-invariant cascaded forest," IEEE Access, vol. 7, pp. 59613–59623, 2019.
[11] X. Wang, W. Zhang, X. Wu, L. Xiao, Y. Qian, and Z. Fang, "Real-time vehicle type classification with deep convolutional neural networks," Journal of Real-Time Image Processing, vol. 16, no. 1, pp. 5–14, 2019.
[12] S. Y. Jo, N. Ahn, Y. Lee, and S. J. Kang, "November. Transfer learning-based vehicle classification," in Proceedings of the 2018 International SoC Design Conference (ISOCC), pp. 127-128, IEEE, Daegu, South Korea, 2018.
[13] J. Kim, J. Kim, G.-J. Jang, and M. Lee, "Fast learning method for convolutional neural networks using extreme learning machine and its application to Lane Detection," Neural Networks, vol. 87, pp. 109–121, 2017.
[14] J. Cao, W. Wang, X. Wang, C. Li, and J. Tang, "October. End-to-End view-aware vehicle classification via progressive CNN learning," in Proceedings of the CCF Chinese Conference on Computer Vision, pp. 729–737, Springer, Singapore, 2017.
[15] L. Jiang, J. Li, L. Zhuo, and Z. Zhu, "August. Robust vehicle classification based on the combination of deep features and handcrafted features," in Proceedings of the 2017 IEEE Trustcom/BigDataSE/ICESS, pp. 859–865, IEEE, Sydney, Australia, 2017.
[16] J. Cao, W. Wang, X. Wang, C. Li, and J. Tang, "October. End-to-End view-aware vehicle classification via progressive CNN learning," in Proceedings of the CCF Chinese Conference on Computer Vision, pp. 729–737, Springer, Singapore, 2017.
[17] X. Liu, W. Liu, T. Mei, and H. Ma, "Provid: progressive and multimodal vehicle reidentification for large-scale urban surveillance," IEEE Transactions on Multimedia, vol. 20, no. 3, pp. 645–658, 2017.
[18] 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," Machine Vision and Applications, vol. 28, no. 7, pp. 793–802, 2017.
[19] L. Jiang, J. Li, L. Zhuo, and Z. Zhu, "August. Robust vehicle classification based on the combination of deep features and handcrafted features," in Proceedings of the 2017 IEEE Trustcom/BigDataSE/ICESS, pp. 859–865, IEEE, Sydney, Australia, 2017.
[20] 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, pp. 2818–2826, Seattle, WA, USA, 2016.
[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778, Seattle, WA, USA, 2016.
[22] C. M. Bautista, C. A. Dy, M. I. Mañalac, R. A. Orbe, and M. Cordel, "May. Convolutional neural network for vehicle detection in low resolution traffic videos," in Proceedings of the 2016 IEEE Region 10 Symposium (TENSYMP), pp. 277–281, IEEE, Bali, Indonesia, 2016.
[23] X. Liu, W. Liu, T. Mei, and H. Ma, "October. A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in Proceedings of the European Conference on Computer Vision, pp. 869–884, Springer, Cham, Switzerland, 2016.Clark, T., Woodley, R., De Halas, D., 1962. Gas-Graphite Systems, in *"Nuclear Graphite"* R. Nightingale, Editor. Academic Press, New York, p. 387.