Classification of Car Body Type using Deep Transfer Learning

Anshul Kumar[1]

[1]B. Tech Student, Dept. of Information Technology, Maharaja Agrasen Institute of Technology, Delhi, India

---

Abstract - *With advancements in computer vision, vehicle classification demonstrates a huge potential to modernize transportation-related systems, Image-based car body type detection being one of them. In this paper, custom-made model based on various predefined transfer learning models is used, which is able to distinguish one type of car from another. It is concluded that after applying some adjustments and hyper-tuning the model, some of the transfer learning-based models perform better than their counterparts. The EfficientNet-based model performs better than all the other transfer learning models under consideration within this research. This study can further be used in fields like autonomous driving, visually impaired environment detection, and traffic management.*

Key Words: Transfer Learning, Car Classification, Convolutional Network, ResNet, DenseNet, EfficientNet

1.INTRODUCTION

With the growth of vehicles on the road, there is a lot of potential in Image-based car classification as it can aid us in the fields like traffic management, surveillance management, and visually impaired environment detection. There are already a lot of different ways to distinguish a car type from another using laser and acoustic sensors. But image-based car body classification has the potential to be the most effective out of the bunch as it is easier to implement and maintain on a larger scale.

Computer vision-based classification is broadly a 2-step process, the first step is to utilize a feature extractor to extract out all the learnable features in the image dataset, the second step is to use a machine learning-based classifier to classify the image into one of the output classes.

Recently, the deep learning-based approach which utilizes Convolutional Neural Network's (CNN's) have been shown to achieve higher accuracy than the classical machine learning-based approach because of their advanced internal architecture [1-2]. Even though the advancements in the graphical processing unit (GPU) have drastically increased image processing ability in modern computers, still deep learning-based approaches face the drawback that it requires large amounts of data to show good performance. To combat this a technique called transfer learning is used, which uses a model trained on different data and uses that pre-trained model for a different task.

Here transfer learning and augmentation of the Images are utilized to combat the data size limitation of deep learning.

With the advancements in computer vision, image and video processing, and pattern recognition various different research has been conducted on vehicle classification or on a sub-vehicle class classification, [3] is able to achieve an accuracy of 97% for vehicle classification here they use traffic vehicle images, AlexNet is utilized as a feature extractor, PCA and LDA are used to process the image and SVM is used as a classifier, [4] is able to use a ResNet_151 as a feature extractor and 3 different classifiers are used to classify different types of truck images.

This research is focused on car body type classification which deals with the classification of cars which is a subclass of vehicles. Given an image of the car from a variety of positions and from a distance, the image is classified into one of the categories i.e., Cab, Convertible, Coupe, Hatchback, Minivan, Sedan, SUV, Van, Wagon, and others. The car body types are very challenging to predict as some of the cars can belong to multiple categories and there aren't many visual differences between categories like convertible and coupe. EfficientNet, DenseNet, and ResNet here are used as feature extractors and a custom-made classifier is used in the model to classify one type of car body type from another.

2. METHODOLOGY

2.1 ResNet

Residual Network (ResNet) [5] proposes that by using a residual network it is possible to build better performing deeper neural networks. The degradation of the gradient is a problem in deep neural networks, it makes it difficult to achieve identity mapping in the network. The skip connection in a ResNet block retains earlier learned information with the newly learned information, this makes it possible to have a memory of previous information even when deeper in the network. The mathematical formulation of the added residual learning units can be expressed as:

$$y = F(x, \{W_i\}) + x$$

Where x and y are, the input and output of the layers respectively; and the function $f(x, \{W_i\})$ is the residual mapping to be learned. Here ResNet_50 is used as a feature extractor.

2.2 DenseNet

Dense Convolutional Network (DenseNet) was developed by Huang, Liu, and Maaten and had the best accuracy on ImageNet in 2017 [6]. DenseNet is made up of ResNet-like skip connections called dense connections, in which each layer is connected to all of the other subsequent layers in the network. So, the features learned by the earlier layer stay intact till the last layer. Due to this property, the gradient is strong throughout the network and the network is able to achieve high accuracy even with a shallower neural network that uses fewer parameters. The mathematical formulation of the dense block can be expressed as:

$$x_l = H_l([x_0, x_1, \dots, x_{l-1}])$$

where $[x_0, x_1, \dots, x_{l-1}]$ refers to the concatenation of the feature- maps produced in layers $0, \dots, l-1$. Here DenseNet_121 and DenseNet_169 are used as feature extractors.

2.3 EfficientNet

EfficientNet [7] is a Convolutional Neural Network that puts an emphasis on the scaling of the network in all 3 dimensions (width, height, resolution) using compound scaling. Unlike conventional Neural Networks which scale only in 1 dimension, EfficientNet scales in all 3 dimensions uniformly. The compound scaling method which is utilized here uses a compound coefficient Φ to uniformly scale the width, depth, and resolution:

$$\text{depth } d = \alpha^{\Phi}$$

$$\text{width } w = \beta^{\Phi}$$

$$\text{resolution } r = \gamma^{\Phi}$$

$$\text{such that } \alpha . \beta^2 . \gamma^2 \approx 2$$

$$\alpha \geq 1, \beta \geq 1, \gamma \geq 1$$

ɸ stands for global scaling factor and it controls how many of the resources are available. α, β, and γ determine how to efficiently assign resources to the depth, width, and resolution respectively.

The MBConv block from MobileNetV2 is used as a fundamental block in EfficientNet [8]. Here EfficientNet_B1 and EfficientNet_B4 are used as feature extractor.

2.4 Dataset

In this paper, the Stanford Car Dataset [9] is used, which contains 8146 labeled images of stationary cars from different angles. The cars belong to 10 different classes: Cab, Convertible, Coupe, Hatchback, Minivan, Sedan, SUV, Van, Wagon, other

Fig -1: Different Car body type examples (a) Cab, (b)Convertible, (c) Coupe, (d) Hatchback, (e) Minivan, (f)Sedan, (g) SUV, (h) Van, (i) Wagon, (j) other

2.5 Augmentation

Deep learning requires huge amount of data to achieve better results, data augmentation is one very easy and effective way to overcome that, it also helps in avoiding overfitting from the network by expanding the dataset through different kinds of transformations. To increase the diversity, 5 different types of data augmentation are used: horizontal flip, rotation, width shift, height shift, zoom.

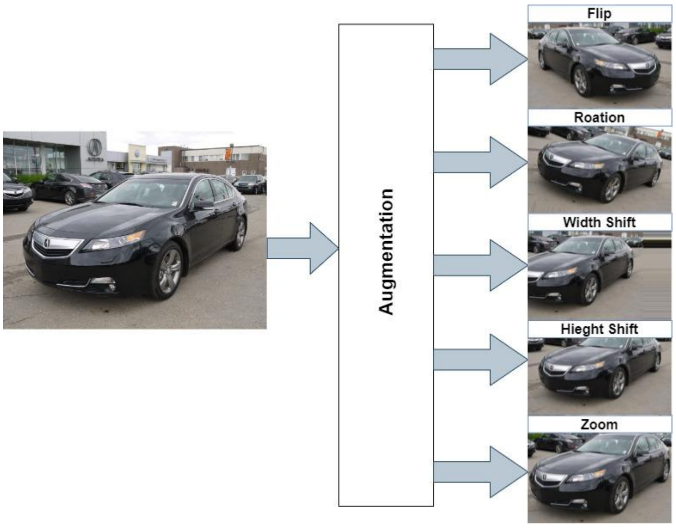

Fig -2 Data Augmentation performed on the dataset

3. ARCHITECTURE

In this paper, a custom-made network is deployed which uses multiple different pre-trained deep transfer learning models as feature extractors. The network's depth depends on the feature extractor being used at the time. The dataset size when compared to the original ImageNet size is very small in comparison so the pre-trained feature extractor is loaded with ImageNet weights and is not trained on the dataset. Each model is tuned to take a 3 channel RGB input of size 250x250. Only the feature extractor part of the pre-trained model is used and not the fully connected classifier layers. For the classifier part of the network, a custom made classifier is developed, it comprises mainly of 3 components, the information from the feature extractor is passed onto a batch normalization layer which accelerates the training by reducing Internal Covariate Shift [10], this helps the network in converging faster towards the result, following the batch normalization is a fully connected layer having 256 neurons and ReLu as the activation function followed by a dropout layer to reduce the gradient degradation problem and to avoid overfitting in the network. Finally, a 10-neuron classification layer is used which uses the softmax function.

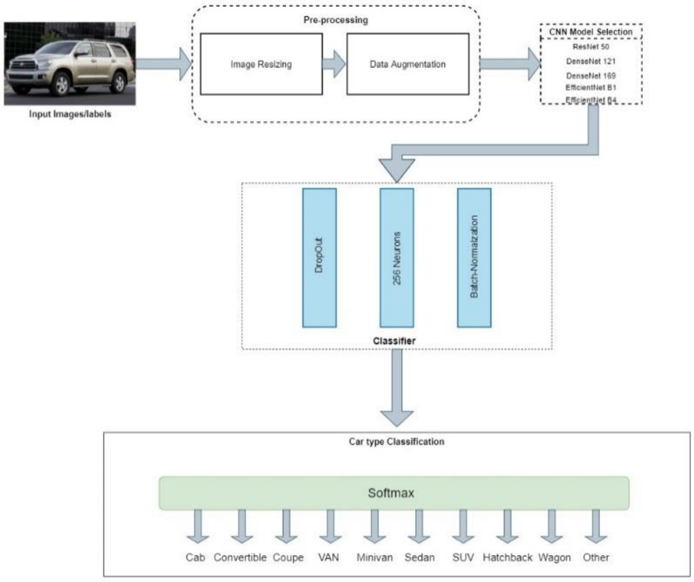

Fig -3 Network architecture

4. EXPERIMENTS AND RESULTS

The proposed class classification method is assessed on a cloud computing setup. The experiments are performed on the heavy computing machine equipped with the Nvidia Tesla P100, 16 GB DDR5 GPU.

4.1 Training of the classification system

The whole training process comprises of 3 main sections:

(i) Data Preprocessing

(ii) Training

(iii) Evaluation

In the first step, the images from the dataset are resized to 250x250 according to the input for the model. Then the training and testing images are randomly split in 80-20 ratio for the train and test split, validation dataset is extracted from the training dataset in 80-20 split as well. TensorFlow 2 libraries and MATLAB R2021a are utilized for the completion of the task (i.e., data preprocessing and organization, training, evaluation, and modification to the model).

The dataset contains 10 different classes of car types but the number of images per class is not the same across all the classes so augmentation is used to get a more uniform dataset. Before augmentation, there are 8146 labeled images after augmentation 10000 labeled images are there in the training dataset.

ResNet, DenseNet, EfficientNet are then used as the feature extractor for the classification model. The training of these was performed on the TensorFlow framework. Adam [11] optimizer is in use the hyperparameters can be found in table 1, the model is trained on a batch size of 30.

Cross-entropy is used as a loss function to accumulate the loss during the training process, validation is performed after each epoch to evaluate the performance while training the network.

Table -1: Hyperparameters of the network

Hyperparameters	
Optimizer	Adam
Learning Rate	0.001
ϵ	0.0000001
Loss	Cross-entropy
Epochs	45

4.2 Evaluation

When training on the augmented dataset containing 10 different car body types as classes all 5 of transfer learning-based models were able to perform decently on the dataset when it comes to training, it is observed from Fig 3 that both models using EfficientNet as its feature extractor learn the fastest, they also show the best train accuracy as seen in chart-1. Whereas ResNet based model performs the worst when compared to DenseNet and EfficientNet based models when it comes to training accuracy and the speed of learning.

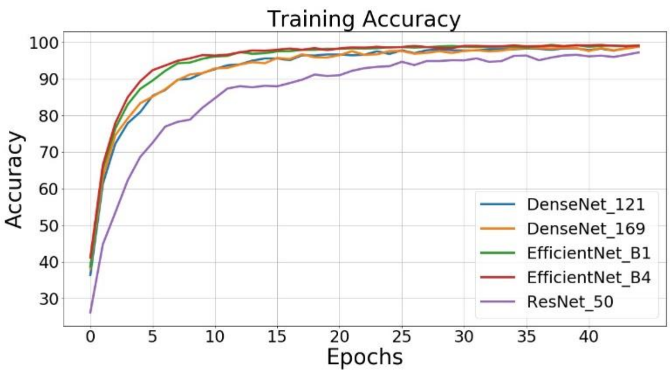

Chart-1: Network training accuracy graph

As per table 2 on the dataset out of all the different models, EfficientNet_B4 based model performs the best achieving the test accuracy of 87.81, and the ResNet_50 based model performs the worst out of all its counterparts achieving a test accuracy of 74.21.

Table -2: Test Accuracy of model for different feature extractors

Feature extractor	Accuracy
EfficientNet_B4	87.73
EfficientNet_B1	87.08
DenseNet_169	85.16
DenseNet_121	84.87
ResNet_50	74.21

5. CONCLUSION

In this paper, CNN-based transfer learning models are proposed to improve the performance of Intelligent Car Body Type Detection in autonomous driving. A dataset containing 10000 different images belonging to 10 different classes is trained on. Augmentation is used on the dataset to get a more uniform dataset. Sate of art Transfer learning models like ResNet, EfficientNet, and DenseNet are used in cooperation with custom fine-tuned classifiers. It is observed that EfficientNet_B4 based model performs the best with an impressive test score of 87.81. This study can further be extended by using a more fine-grained model to further aid the effectiveness of the proposed method in computer vision-related tasks.

REFERENCES

[1] R. Girshick, "Fast r-cnn," arXiv preprint arXiv:1504.08083, 2015.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in Advances in neural

[3]. Zhou, Yiren & Cheung, Ngai-Man. (2016). Vehicle Classification using Transferable Deep Neural Network Features.

[4] R. V. Nezafat, B. Salahshour and M. Cetin, "Classification of truck body types using a deep transfer learning approach," 2018 21st International Conference on Intelligent Transportation Systems (ITSC), 2018, pp. 3144-3149, doi: 10.1109/ITSC.2018.8569514.

[5] Kaiming He and Xiangyu Zhang and Shaoqing Ren and Jian Sun (2015). Deep Residual Learning for Image Recognition. CoRR, abs/1512.03385.

[6] Gao Huang and Zhuang Liu and Kilian Q. Weinberger (2016). Densely Connected Convolutional Networks. CoRR, abs/1608.06993.

[7]. Mingxing Tan and Quoc V. Le (2019). EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. CoRR, abs/1905.11946.

[8] Mark Sandler and Andrew G. Howard and Menglong Zhu and Andrey Zhmoginov and Liang-Chieh Chen (2018). Inverted Residuals and Linear Bottlenecks: Mobile Networks for Classification, Detection and Segmentation. CoRR, abs/1801.04381.

[9] Jonathan Krause, Michael Stark, Jia Deng, & Li Fei-Fei (2013). 3D Object Representations for Fine-Grained Categorization. In 4th International IEEE Workshop on 3D Representation and Recognition (3dRR-13).

[10] Sergey Ioffe, & Christian Szegedy. (2015). Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.

[11]. Diederik P. Kingma, & Jimmy Ba. (2017). Adam: A Method for Stochastic Optimization.

[12] J. Chang, L. Wang, G. Meng, S. Xiang, and C. Pan, “Vision-based occlusion handling and vehicle classification for traffic surveillance systems,” IEEE Intelligent Transportation Systems Magazine, vol. 10, no. 2, pp. 80–92, 2018.

[13] J. Cai, J. Deng, M. S. Khokhar, and M. U. Aftab, “December. Vehicle classification based on deep convolutional neural networks model for traffic surveillance systems,” in Pro-ceedings of the 2018 15th International Computer Conferenceon Wavelet Active Media Technology and Information Pro-cessing (ICCWAMTIP), pp. 224–227, IEEE, 2018.