Project Report

Team Member Names:

Ryan Elson

Project Title:

Classifying Car Images with Machine Learning

Contents

Problem Statement

Image classification is an important field that is broadly used across various industries. Whether looking at MRIs to determine presence of a medical issue, analyzing remote sensing data to determine what type of ground coverage a satellite is viewing, training a self-driving car, or even looking at products on an assembly line to locate defects, image classification is at the heart of these applications.

With this in mind, I decided that it would be interesting to build a model(s) to classify images of cars. The model will be used to try and predict the car (make, model, year) from a test image dataset as well as predicting the body type of car (e.g., sedan, SUV, truck), and whether the car's miles per gallon (MPG) or miles per gallon equivalent (MPGe) is high or low, a non-visual characteristic that may be influenced or correlated with the car's physical form. Although there are datasets used to predict MPG based on car features (1), such as engine type, model year, and weight, I was unable to find any instances of image classification models being used to try and classify MPG directly. Intuitively, it seems like this approach should be successful, and this is an interesting and novel use of image classification.

Data Source

For this analysis, I located a dataset with images of cars that was originally used in a paper from 2013 entitled, "3D Object Representations for Fine-Grained Categorization," by Krause, et al (2). The entire dataset has 16,185 images, roughly split in half between training and testing subsets. There were 8144 training images and 8041 test images.

There are additional MAT and CSV files containing label/class information for the type of car (make, model, and year) and "x" and "y" coordinates for cropping the images.

Some challenges of this dataset are the size of the images and number of files, as well as the large number of classes. There are 196 different classes of car, some of which are extremely similar, which makes achieving high accuracy difficult.

The dataset was available through the Stanford Artificial Intelligence (AI) department's website (3) and through Kaggle (4).

Methodology

Once the dataset was downloaded, the first step was working with the associated files. Two separate CSV files contained training and test subset details. These files contained X, Y coordinates for cropping the car images. However, the test dataset details did not contain the true classes. In addition, classes for the training dataset were numeric and associated make, model, year details were not included. Data from CSV files was as shown below.

Table 1: Raw Training Dataset CSV Details

	x1	y1	x2	y2	Class	image
0	39	116	569	375	14	00001.jpg
1	36	116	868	587	3	00002.jpg
2	85	109	601	381	91	00003.jpg

3	621	393	1484	1096	134	00004.jpg
4	14	36	133	99	106	00005.jpg
5	259	289	515	416	123	00006.jpg

Table 2: Raw Test Dataset CSV Details

	x1	y1	x2	y2	image
0	30	52	246	147	00001.jpg
1	100	19	576	203	00002.jpg
2	51	105	968	659	00003.jpg
3	67	84	581	407	00004.jpg
4	140	151	593	339	00005.jpg
5	20	77	420	301	00006.jpg

To obtain the testing subset's classifications, I read the details from one of the MAT files and created my own test data CSV which mirrored the provided training data CSV. This was required so I could evaluate model accuracy on the test data.

Using a different MAT file, I read in the class details so I could associate the numeric classifications with the make, model, and year of the cars. This gave me a two-column CSV, which I manually updated by adding my own additional classifications, body type and a binary MPG classification.

I assumed that creating additional classifications for body type and MPG would be straightforward, but this proved not to be the case.

For body type, depending on the website or publication you read, there are different vehicle segments. Car and Driver and Kelly Blue Book (KBB) both categorize vehicles into eight segments (5, 6) while users can search ten body type segments on Cars.com (7).

Car and Driver	**KBB.com**	**Cars.com**
Convertible	Convertible	Cargo Van
Coupe	Coupe	Convertible
Hatchback	Hatchback	Coupe
Pickup Truck	Sedan	Hatchback
Sedan	SUV/Crossover	Minivan
Station Wagon	Truck	Passenger Van
SUV	Van/Minivan	Pickup Truck
Van	Wagon	Sedan
		SUV
		Wagon

I decided to use the eight categories from Car and Driver for creating classifications since the segments matched well with KBB. These eight segments also matched up with the ten segments from Cars.com if all three types of vans were combined into a single segment.

After deciding on the body type segments to be used, decisions still had to be made on what category some vehicles should be classified in. Details on the selection of body type segments and on how vehicles were classified are in the Appendix.

Next, I created an MPG classification, first gathering MPGs and then creating a binary classifier for vehicles that were above or below the median MPG. This was not as easy as anticipated since vehicles are often available with different size engines which can affect fuel efficiency. I decided to use the EPA's Combined City/Hwy MPG values (8), and for vehicles with more than one entry (i.e., for different engines, transmissions) I averaged the values and rounded to the nearest tenth. If a vehicle version was clearly stated (e.g., hybrid), then I only used that version's entries. If not stated, then I averaged all values. I did not include E85 gas ratings.

For battery electric vehicles, I used MPGe. For plug-in hybrid vehicles, I used the MPGe for electric driving and MPG for gas driving to calculate an average overall rate.

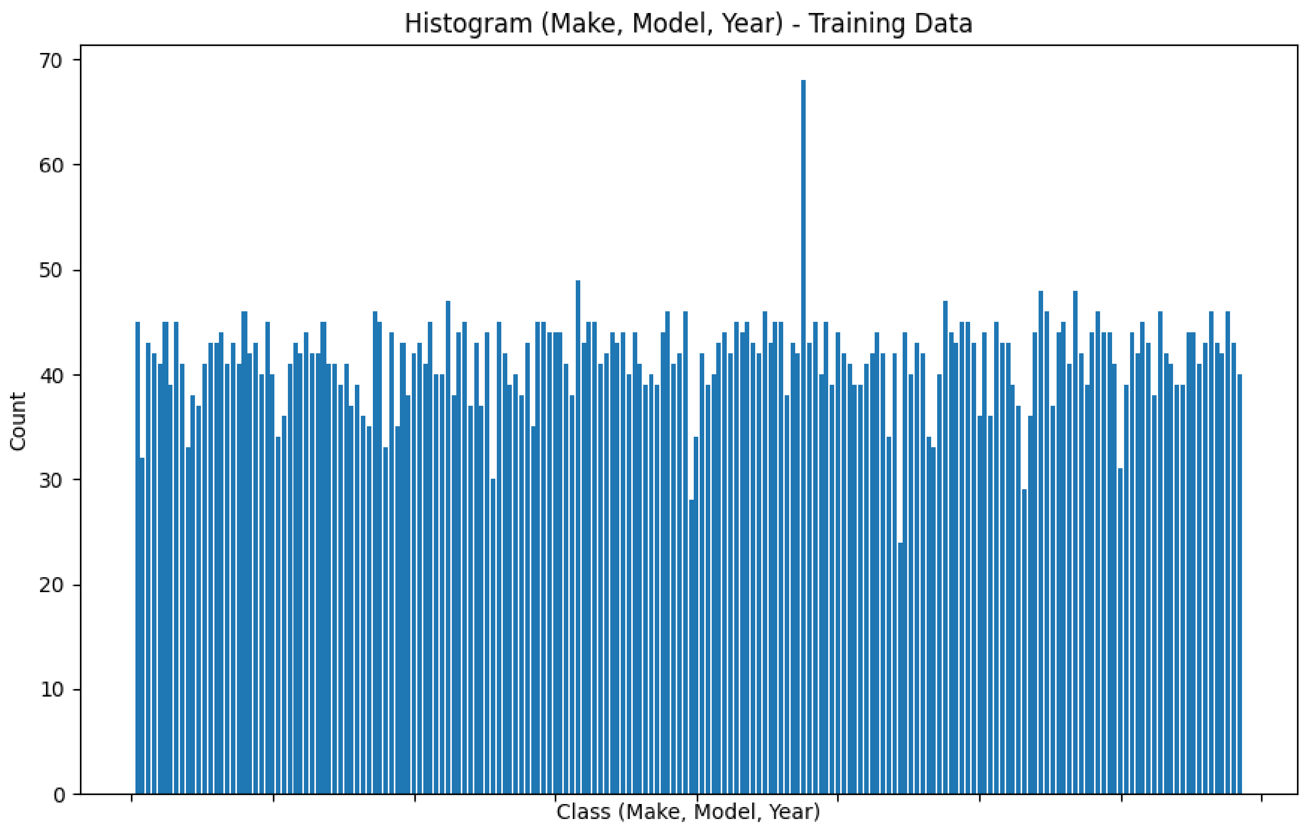
Some vehicles did not show up on the EPA's website because heavy vehicles aren't regulated like smaller vehicles (9), and they do not have to report fuel economy information to the EPA. I searched for these vehicles manually to determine an MPG value. See the Appendix for details.

After all data was gathered, I calculated the median MPG as 18.7. This was the median of the 196 classes not the training dataset since the training dataset's distribution could have been different from the test dataset. I used the median instead of the mean so high MPGe values like for the 2012 Tesla or Nissan Leaf wouldn't skew the results.

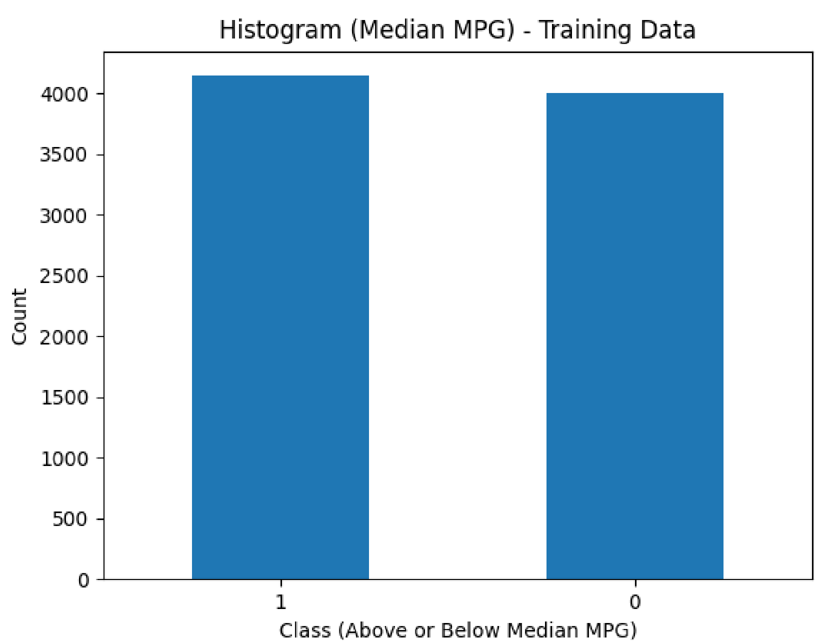
I considered adding a safety rating from the NHSTA as one additional classification measure, but when I searched the website, I found that some of the vehicles in the dataset did not have ratings or did not have overall ratings. As a result, I decided not to attempt to classify based on incomplete safety ratings.

Exploring the data, a few items to note are that the earliest model year is 1991 and the most recent is 2012. Vehicle shapes have changed over this period which will make accurate classification challenging.

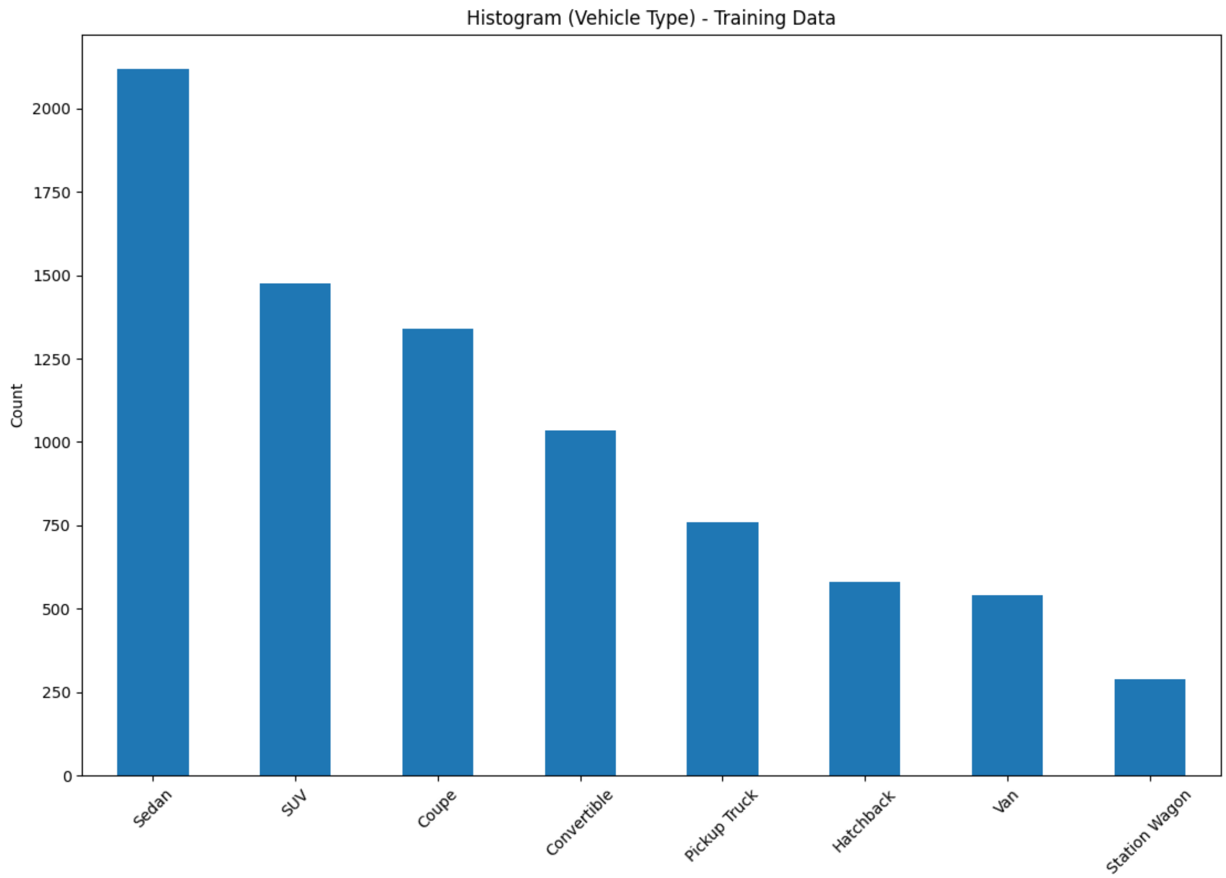
To check if classes are well balanced, I created histograms for all three criteria that I am building models for. Distributions are as shown below.



Histogram (Make, Model, Year) - Training Data
70
60
50
40
30
20
10
0
Count
Class (Make, Model, Year)



Histogram (Median MPG) - Training Data
4000
3500
3000
2500
2000
1500
1000
500
0
Count
1
0
Class (Above or Below Median MPG)



The histograms for “Make, Model, Year” and “Above or Below Median MPG” are both relatively well balanced, though there are a few underrepresented classes and one obviously overrepresented class in the “Make, Model, Year” plot. The histogram for body type shows a much larger imbalance.

To account for these imbalances, I used a cost-sensitive approach (sklearn’s class_weight method) when fitting my model. This approach adjusts relative weights of observations.

Finally, to prepare the images for analysis, I used the “x” and “y” coordinates from the MAT and CSV files to crop the images and then resized them so they all had the same dimensions. All processed images were saved to new folders so they could be used when building and testing my model.

Cropping the images gets rid of the surrounding area that isn’t of interest in the analysis and resizing them is important so the model can run efficiently and so the model can process the data.

To build and test my model, I opted to use Keras, an API built on top of TensorFlow. Keras uses convolutional neural networks (CNN) and can be used to build different neural network models with different numbers of hidden layers and parameters. There are various references and tutorials online which I used to learn how to perform image classification with Keras; e.g., (10, 11).

From the ISYE 6740 Module 11 lecture (12), the structure of a neural network may be shown as an input layer, some number of hidden layers, and eventually the output layer.

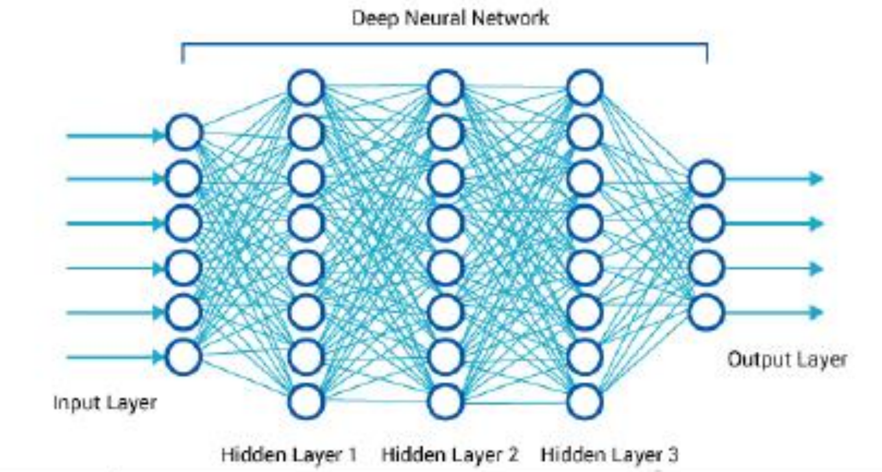


Figure 1: Neural Network depiction (from ISYE 6740, Module 11)

To tune the models, I used the Keras Tuner package and tried different learning rates as well as different numbers of filters for some of the layers. This is an important step since tuning hyperparameters will generally improve a model's performance.

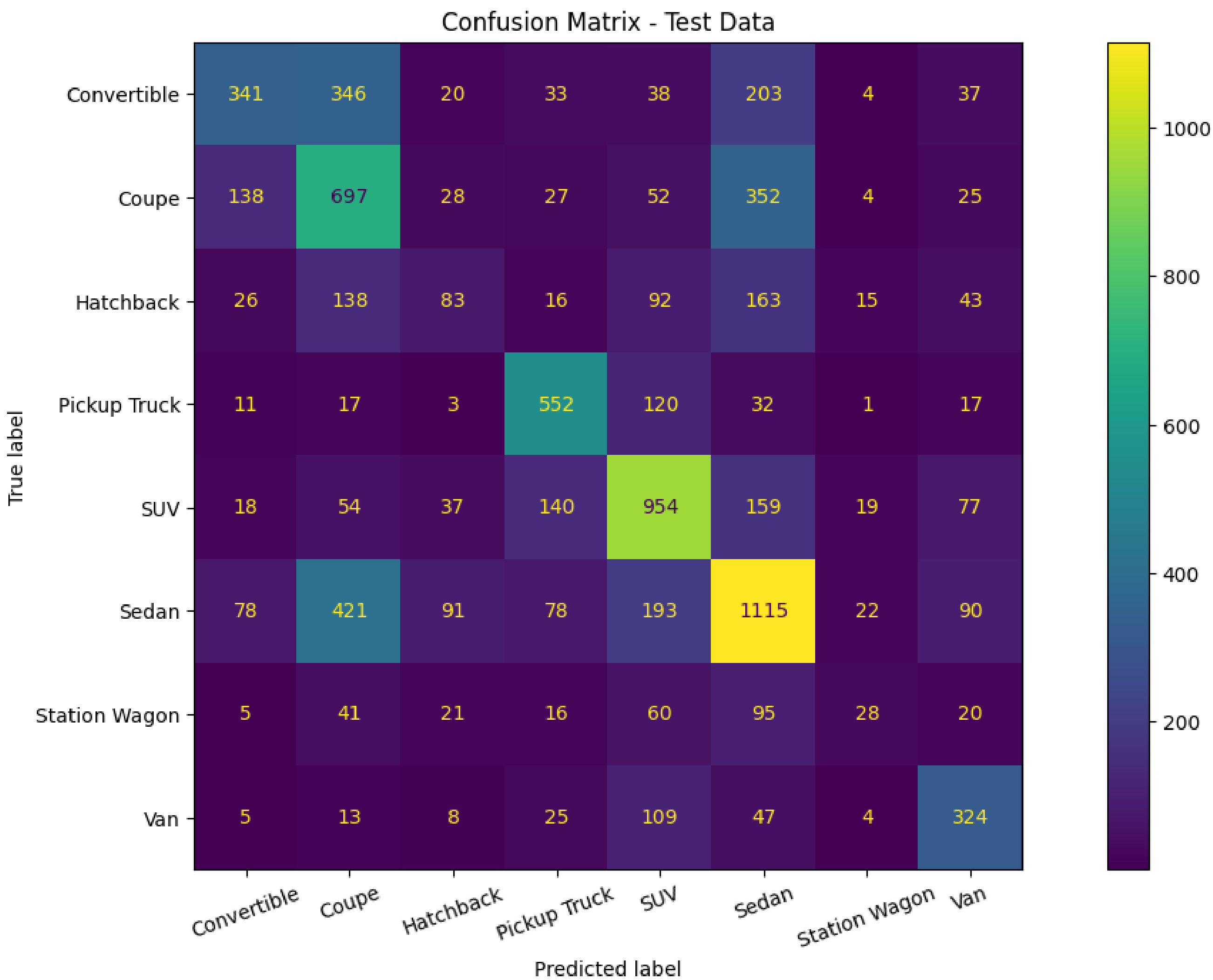
When building and tuning all three separate models, to avoid overfitting I used the Keras EarlyStopping method to stop when the validation loss was no longer decreasing (with a patience delay to ensure this was the case). At that point, even if the training accuracy improved, that would indicate overfitting of the model.

I will select the model with the best tuning parameters that leads to the smallest validation loss for each different classification. For the models with fewer classes (body type and Median MPG), I will also look at the confusion matrix and classification reports (i.e., precision, recall, F1 score). The final accuracy rates will be those attained when the validation loss was the smallest. This is considered the point at which I start overfitting the model when the training loss continues to decrease but the validation loss flattens out or begins to increase. Although validation accuracy can potentially be slightly higher even when validation loss is not the smallest, this may include random effects so I will report accuracy when the validation loss is smallest. Validation loss plots will be in the Appendix.

Evaluation and Final Results

Body Type Classification
Some body types are very similar to each other, which made this classification challenging. Regardless, the model was able to predict body type across eight classes with 50.91% accuracy.

The confusion matrix for test image classification is presented below along with classification statistics, and a summary of common misclassifications. Common Misclassifications are shown in red if that class was predicted more frequently than the true label.



	precision	recall	f1-score	support
Convertible	0.55	0.33	0.41	1022
Coupe	0.40	0.53	0.46	1323
Hatchback	0.29	0.14	0.19	576
Pickup Truck	0.62	0.73	0.67	753
SUV	0.59	0.65	0.62	1458
Sedan	0.51	0.53	0.52	2088
Station Wagon	0.29	0.10	0.15	286
Van	0.51	0.61	0.55	535
accuracy			0.51	8041
macro avg	0.47	0.45	0.45	8041
weighted avg	0.50	0.51	0.50	8041

	Common Misclassifications
Convertible	**Coupe**, Sedan
Coupe	Sedan, Convertible
Hatchback	**Sedan, Coupe, SUV**
Pickup Truck	SUV
SUV	Sedan, Pickup Truck
Sedan	Coupe, SUV
Station Wagon	**Sedan, SUV, Coupe**
Van	SUV

The main hurdle to achieving high accuracy is that many vehicle body types are visually similar to other vehicle body types. For example, sedans and coupes can be very similar and are getting harder to tell apart (5).

Figure 1: Coupe vs Sedan, https://www.caranddriver.com/features/a28411772/sedan-vs-coupe/

In addition, many hatchbacks and station wagons (which were the body types with the worst classification accuracy) look similar to sedans and SUVs. Convertibles also proved challenging when images of the convertibles showed them with the top up. In these images, the convertibles often looked more like sedans or coupes. Finally, the angle an image was taken at could affect prediction accuracy. Images taken directly from the front or rear of a car were challenging. Hard angles and unusual lighting or background conditions also seemed to contribute to incorrect predictions. A selection of misclassified images exhibiting these types of problems is shown below. These images are displayed prior to cropping and resizing although the model used the cropped and resized images for analysis.

Figure 2: Convertible (prediction: Sedan)

Figure 3: Convertible (prediction: Coupe)

Figure 4: Hatchback (prediction: Van)

Figure 5: Hatchback (prediction: Van)

Figure 6: Hatchback (prediction: SUV)

Figure 7: Pickup Truck (prediction: Sedan)

Figure 8: SUV (prediction: Sedan)

Figure 9: Sedan (prediction: Coupe)

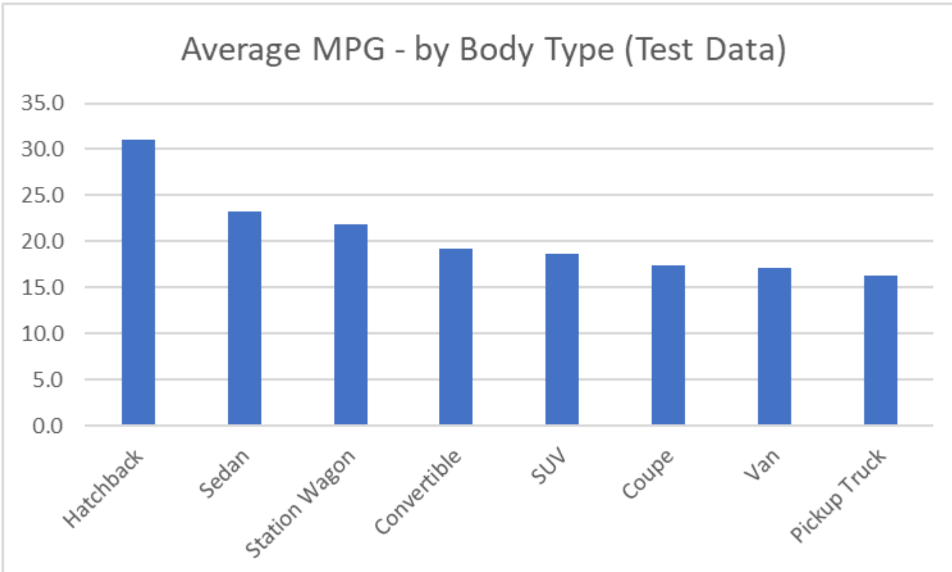
Figure 10: Station Wagon (prediction: SUV)

Figure 11: Station Wagon (prediction: Sedan)

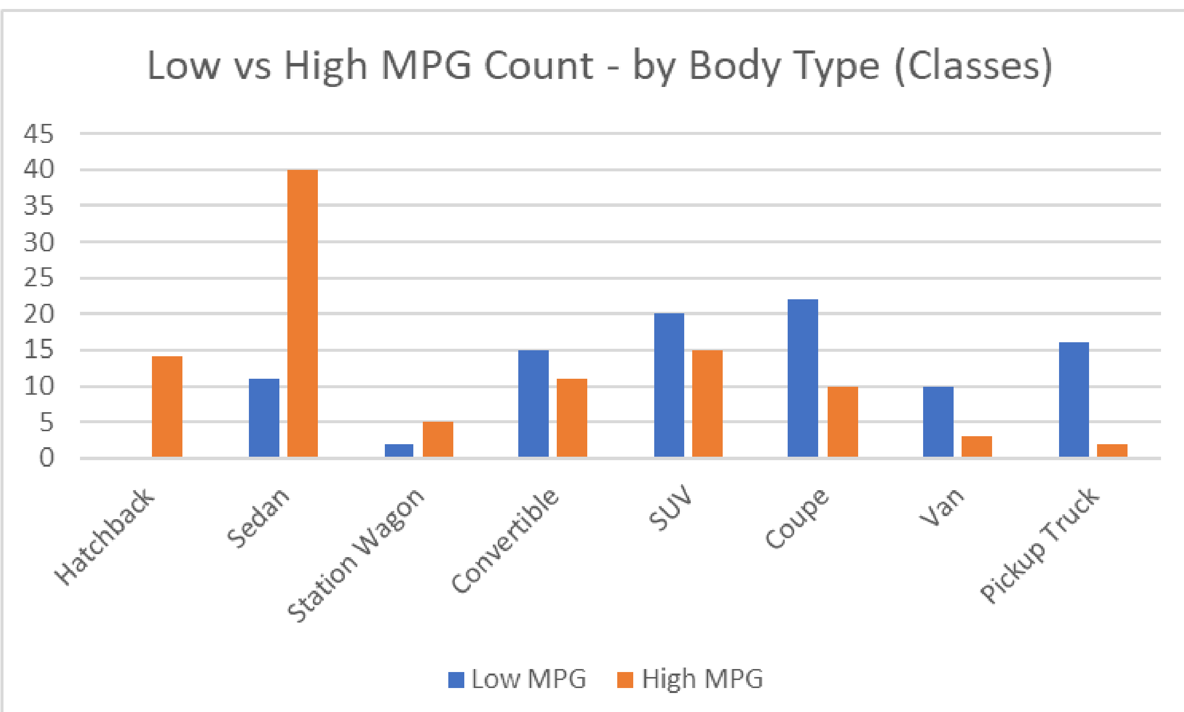
Figure 12: Van (prediction: SUV)

Figure 13: Van (prediction: SUV)

Median MPG
Different categories of car often have higher or lower average fuel economy, so I plotted the average MPG across the different vehicle classes to understand the data. As shown, hatchbacks have the highest average fuel economy and pickups have the worst. Since the different body types vary in appearance, intuitively it seems like using images to try and predict Low vs High MPG (below or above Median MPG) should yield a model with good prediction accuracy.



However, when looking at the body type classes separated by Low or High MPG the picture is less clear.

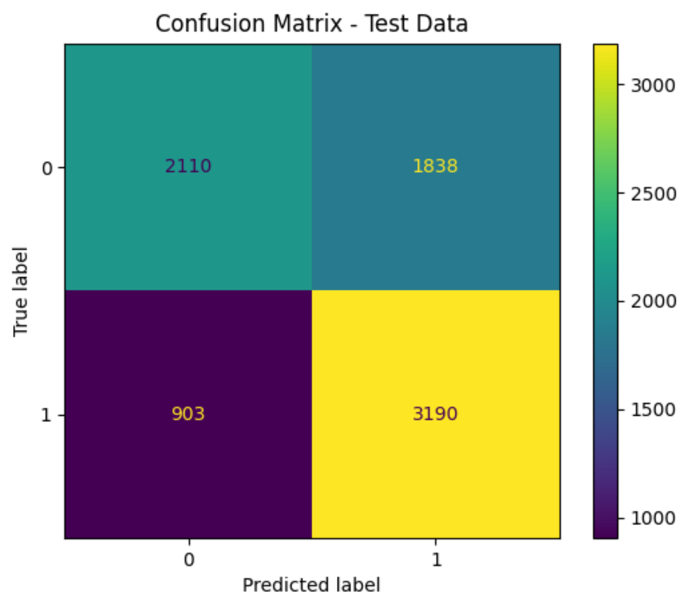


Although all hatchbacks had good fuel economy, and most vans and pickups had poor fuel economy, some car types (convertible, SUV) had many examples that fell into both Low and High MPG classes.

This makes differentiating between vehicles that are visually similar but fall into different MPG classes challenging and will negatively affect prediction accuracy. In addition, and as discussed previously, sedans and coupes can appear similar and it's evident from this plot that sedans are often classified as High MPG while coupes are often classified as Low MPG.

With this in mind, the results of my analysis were positive. After my model stopped when the validation loss was minimized, the test dataset was predicted with 65.91% accuracy. This is a significant result since the test dataset was well balanced with a 49.10/50.90% Low/High MPG split and I used class weighting to further try and account for any imbalance.

As shown by the confusion matrix and associated metrics, the model is somewhat better at predicting high MPG vehicles (vehicles above the median MPG) correctly.



	precision	recall	f1-score	support
0 (Low MPG)	0.70	0.53	0.61	3948
1 (High MPG)	0.63	0.78	0.70	4093
accuracy			0.66	8041
macro avg	0.67	0.66	0.65	8041
weighted avg	0.67	0.66	0.65	8041

Make, Model, Year
Predicting the correct Make, Model, Year classification proved to be the most challenging task. This was expected and makes sense considering there are 196 classes and many of the vehicles and images are very similar to each other.

The best model's overall accuracy rate was 14.69% which initially didn't seem very good to me. However, random chance would dictate an accuracy rate of approximately 0.51% so the model is an improvement of more than 28x.

For the test dataset, it is also informative to look at a random selection of some predicted classes versus actual "Make, Model, Year" classes. As shown in the table below, there are many misclassifications that seem quite close to the true class and if we were to score accuracy not only on the prediction itself but on the actual class being one of the top five predictions then the accuracy would likely improve a good deal.

PREDICTION	**ACTUAL**	CORRECT
Lamborghini Diablo Coupe 2001	Lamborghini Gallardo LP 570-4 Superleggera 2012	
Spyker C8 Coupe 2009	Spyker C8 Convertible 2009	
Lamborghini Reventon Coupe 2008	Bugatti Veyron 16.4 Convertible 2009	
Chevrolet Cobalt SS 2010	Eagle Talon Hatchback 1998	
Cadillac Escalade EXT Crew Cab 2007	Spyker C8 Coupe 2009	
Mercedes-Benz SL-Class Coupe 2009	Aston Martin Virage Convertible 2012	
Toyota 4Runner SUV 2012	BMW X3 SUV 2012	
Dodge Durango SUV 2007	Chevrolet Express Van 2007	
Chevrolet Corvette Convertible 2012	Ford Mustang Convertible 2007	
Mercedes-Benz Sprinter Van 2012	Dodge Sprinter Cargo Van 2009	
Chevrolet Corvette Convertible 2012	Bentley Continental GT Coupe 2007	
Jaguar XK XKR 2012	McLaren MP4-12C Coupe 2012	
Jeep Liberty SUV 2012	Jeep Liberty SUV 2012	YES
Audi TT RS Coupe 2012	Audi TT RS Coupe 2012	YES
Acura TSX Sedan 2012	Fisker Karma Sedan 2012	
Cadillac Escalade EXT Crew Cab 2007	Cadillac SRX SUV 2012	
GMC Acadia SUV 2012	Lincoln Town Car Sedan 2011	
Mercedes-Benz Sprinter Van 2012	Dodge Sprinter Cargo Van 2009	
Chevrolet TrailBlazer SS 2009	GMC Canyon Extended Cab 2012	
Geo Metro Convertible 1993	Geo Metro Convertible 1993	YES
Acura TSX Sedan 2012	BMW 6 Series Convertible 2007	
Bentley Mulsanne Sedan 2011	Bentley Mulsanne Sedan 2011	YES

A few examples of predictions that intuitively seem close to the correct classification (but are nonetheless incorrect) are a Lamborghini Gallardo being misclassified as a Lamborghini Diablo or a Spyker C8 Convertible being misclassified as a Spyker C8 Coupe. There are numerous other examples including a Dodge sprinter van being misclassified as a Mercedes-Benz sprinter van, a

Chrysler Express Cargo van being misclassified as a Chrysler Express Van, and a Chevy Silverado 1500 Regular Cab being misclassified as a Chevy Silverado 1500 Hybrid Crew Cab.

When investigating "Make, Model, Year" classifications more closely, it's easy to compare the predictions to the actual classes and find evidence of these intuitively close classifications. Even though the model was only trying to label the images correctly into 1 of 196 "Make, Model, Year" classes, there are an additional 2872 misclassifications (35.72%) that were still correct for either the car's body type or the vehicle's make even though incorrect for the exact "Make, Model, Year."

Conclusion

In conclusion, I was able to use image classification to predict a non-visual MPG classification with 65.91% accuracy, body type classification with 50.91% accuracy, and "Make, Model, Year" classification with 14.69% accuracy. These are all significant results and show the power of machine learning for image classification.

Future work may include additional image classification tasks such as training a model on certain views (e.g., front, rear, side, isometric) of vehicles rather than including all views in a single class. It should also be possible to identify the symbol of the vehicle's make and use it to improve prediction accuracy. For example, it's unlikely that a vehicle identified as a Buggati will be a station wagon.

The main lessons learned were how to perform image classification using Keras and how important tuning hyperparameters can be. Initially, being new to image classification using Keras and Tensorflow, I simply tried to build a functional model and start understanding the package. For this reason, I started by hard-coding the number of filters and learning rates in my models and only achieved a "Make, Model, Year" accuracy of 4.07%. Tuning hyperparameters brought that accuracy rate up to 14.69%, so the improvement was substantial. The other models' accuracy rates also improved. I also tried different Image Generator parameters like rotation range and whether to flip images in order to improve each model's performance.

Appendix

Additional Classification Details – Body Type
In addition to the body type segments found on Car and Driver, KBB.com, and Cars.com, other companies and websites have their own groupings. For example, car rental companies often let customers search by categories like economy, small, medium, intermediate, and luxury, and there are 19 EPA class sizes for 2022 (8). Some websites even have different categories on different webpages within the site. Cars.com's research page has categories like Green Car/Hybrid, Sports Car, and Minivan/Van, which are different from the ten body type categories previously discussed. Regardless, the eight selected categories seem intuitive and show consistency across several websites so were selected for the analysis.

After deciding on the categories for body type, classification decisions had to be made for each car. I decided that if one of the eight categories (or a derivation or clear indication thereof: e.g., pickup truck, truck, extended cab, crew cab) was specifically in the title then I would use that classification. This meant that "Aston Martin V8 Vantage Convertible 2012" was categorized as a convertible, even though a search on Cars.com for 2012 Aston Martins returned matching models as "V8 Vantage" and "V8 Vantage S," both of which were listed as coupes. Digging deeper, the car could be purchased as a coupe or convertible. One car had two classifications in the title (Rolls-Royce Phantom Drophead Coupe Convertible 2012), and I used the classification that seemed more distinct, so this vehicle was classified as a convertible. Manual inspection of associated images confirmed this classification.

Of the 196 "make, model, year" classes, 14 were not yet assigned a body type at this point, so I used Cars.com to search for the body type for the remaining vehicles. A few of these vehicles were available for purchase in more than one body type (sedan or coupe), so I used the detailed class description to try to pick the correct classification and looked at a few images to verify.

Additional Classification Details – MPG

A number of vehicles did not have MPG values on the EPA's website, either because they were heavy and did not fall under the same EPA regulations or for another unknown reason. How these vehicles were assigned MPG values is discussed below.

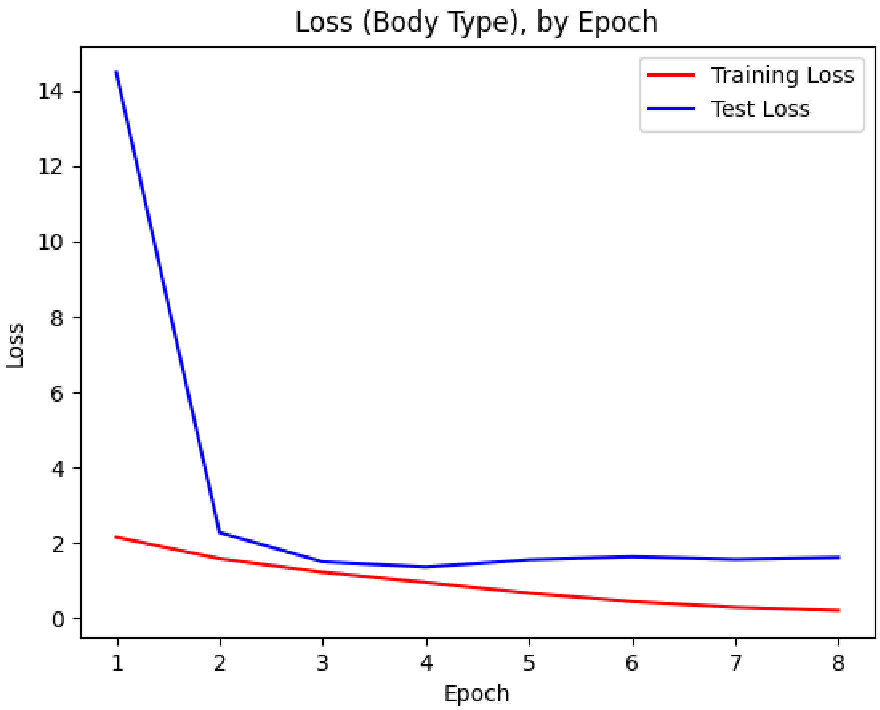
The Bugatti Veyron was not listed on the EPA's website for 2009, so I used the 2010 MPG rating.

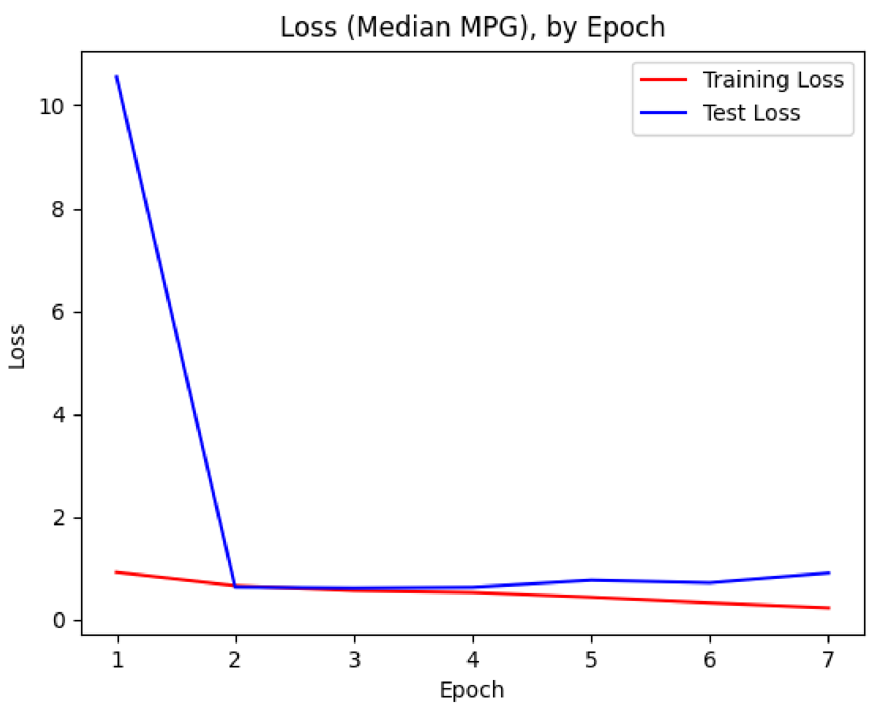
The 2009 and 2010 Dodge 3500 pickup wasn't listed either, so I used the 1500 values instead even though the MPG should be worse for the 3500. For the 2012 Chevy Silverado 2500, I used the 2012 Chevy Silverado 15/1500 values. For the 2012 Ford F450, I used the F150 MPG. Again, actual MPG values were likely worse than those used.

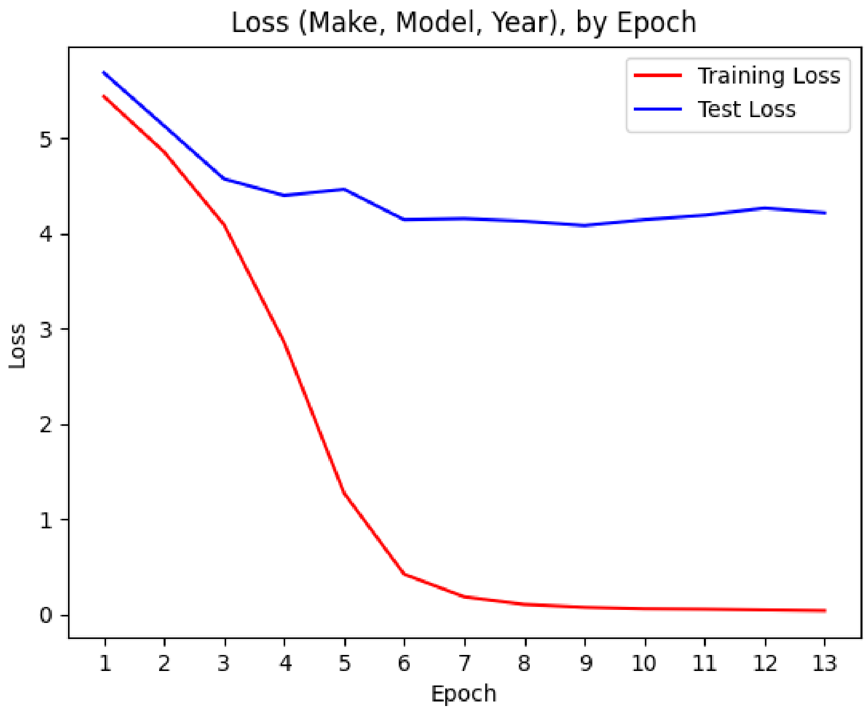
The 2009 Hummer H2 wasn't listed, so I used the H3T since it was the only Hummer pickup listed from that year.

I used values for the 2012 Mercedes Benz sprinter (13), the 2009 Dodge Sprinter Van (14), the 2012 Nissan NV passenger van (15), and the 2000 Hummer (16) from Motortrend.

Loss Plots


Loss (Body Type), by Epoch
Training Loss
Test Loss
Loss
Epoch
0
2
4
6
8
10
12
14
1
2
3
4
5
6
7
8


Loss (Median MPG), by Epoch
Training Loss
Test Loss
Loss
Epoch
0
2
4
6
8
10
1
2
3
4
5
6
7



References

1. UCI.edu. "UCI Machine Learning Repository: Auto MPG Data Set," 2022. https://archive.ics.uci.edu/ml/datasets/auto+mpg

2. "3D Object Representations for Fine-Grained Categorization", Jonathan Krause, Michael Stark, Jia Deng, Li Fei-Fei, 4th IEEE Workshop on 3D Representation and Recognition, at ICCV 2013 (3dRR-13). Sydney, Australia. Dec. 8, 2013.

3. Krause, Jonathan. "Cars Dataset." Stanford Artificial Intelligence Laboratory, https://ai.stanford.edu/~jkrause/cars/car_dataset.html.

4. Li, Jessica. "Stanford Cars Dataset." Kaggle, 5 June 2018, https://www.kaggle.com/datasets/jessicali9530/stanford-cars-dataset.

5. Stafford, Eric. "Sedan vs. Coupe: How Different Are They?" Car and Driver. Car and Driver, July 24, 2019. https://www.caranddriver.com/features/a28411772/sedan-vs-coupe/.

6. KBB.com. "Certified 2018 Honda Accord LX," 2018. https://www.kbb.com/cars-for-sale/all.

7. Cars.com. "Cars.com," 2022. https://www.cars.com/.

8. Fueleconomy.gov. "Search for Model Year 2022 Vehicles by EPA Size Class," 2022. https://www.fueleconomy.gov/feg/byclass/2022ClassList.shtml.

9. Edelstein, Stephen. "You'll Never Know the EPA Ratings of Heavy-Duty Pickup Trucks: Here's Why." Green Car Reports, August 11, 2016. https://www.greencarreports.com/news/1105494_youll-never-know-the-epa-ratings-of-heavy-duty-pickup-trucks-heres-why.

10. Eijaz Allibhai. "Building a Convolutional Neural Network (CNN) in Keras." Medium. Towards Data Science, October 16, 2018. https://towardsdatascience.com/building-a-convolutional-neural-network-cnn-in-keras-329fbbadc5f5

11. Vijayabhaskar J. "Tutorial on Keras Flow_from_dataframe - Vijayabhaskar J - Medium." Medium. Medium, September 21, 2018. https://vijayabhaskar96.medium.com/tutorial-on-keras-flow-from-dataframe-1fd4493d237c.

12. Xie, Y. "Module 11: Neural Networks." Georgia Tech, ISYE 6740. Spring 2022.

13. Sempson, Duane. "2012 Mercedes-Benz Sprinter 2500 Bluetec Update 1 - Motor Trend." MotorTrend. MotorTrend, February 2014. https://www.motortrend.com/reviews/2012-mercedes-benz-sprinter-2500-bluetec-update-1/.

14. Evans, Scott. "2009 Dodge Sprinter 2500 Passenger van First Drive - Motor Trend." MotorTrend. MotorTrend, October 13, 2009. https://www.motortrend.com/reviews/2009-dodge-sprinter-2500-passenger-van-drive/.

15. Sempson, Duane. "2012 Nissan NV 2500HD Long-Term Verdict - Motor Trend." MotorTrend. MotorTrend, March 20, 2013. https://www.motortrend.com/reviews/2012-nissan-nv-2500hd-long-term-verdict/.

16. Walton, Chris. "Comparison of SUVs - Motor Trend." MotorTrend. MotorTrend, November 2, 1999. https://www.motortrend.com/features/2000-am-general-hummer/.

17. Stewart, Matthew. "Guide to Classification on Imbalanced Datasets - towards Data Science." Medium. Towards Data Science, July 20, 2020. https://towardsdatascience.com/guide-to-classification-on-imbalanced-datasets-d6653aa5fa23.