Automated Truck Taxonomy Classification Using Deep Convolutional Neural Networks

Abdullah Almutairi[1] · **Pan He[2]** · **Anand Rangarajan[2]** · **Sanjay Ranka[2]**

Abstract

Trucks are the key transporters of freight. The types of commodities and goods mainly determine the right trailer for carrying them. Furthermore, finding the commodities' flow is an important task for transportation agencies in better planning freight infrastructure investments and initiating near-term traffic throughput improvements. In this paper, we propose a fine-grained deep learning based truck classification system that can detect and classify the trucks, tractors, and trailers following the Federal Highway Administration's (FHWA) vehicle schema. We created a large, fine-grained labeled dataset of vehicle images collected from state highways. Experimental results show the high accuracy of our system and visualize the salient features of the trucks that influence classification.

Keywords truck classification · vehicle classification · deep learning · convolutional neural networks · intelligent transportation system

1 Introduction

Intelligent transportation systems (ITS) offer advanced possibilities for modeling and tracking traffic flow, where various innovative services are integrated, to make users better informed, resulting in a smarter, safer, and more coordinated transportation network. With the volume of the traffic growing rapidly, it has attracted increasing attention in the transportation community[1, 31].

Abdullah Almutairi and Pan He contributed equally to the work.

✉ Abdullah Almutairi
abdullah.almutairi@ku.edu.kw

Pan He
pan.he@ufl.edu

Anand Rangarajan
anand@cise.ufl.edu

Sanjay Ranka
ranka@cise.ufl.edu

[1] Department of Information Science, Kuwait University, Al-Shadadiya, Kuwait

[2] Department of Computer and Information Science and Engineering, University of Florida, Gainesville, FL 32611, USA

Freight transportation is the physical process of transporting commodities, merchandise goods and cargo, playing an important role in the area of the ITS[31]. Trucks are largely responsible for transporting freight. According to the study of American Trucking Associations (ATA)[1], the trucking industry continues to dominate the freight transportation industry, in terms of both tonnage and revenue. The study forecasts a long-term positive trend for trucking and the overall freight economy, despite the effects of the COVID-19 pandemic. Overall freight revenues in 2020 totaled $879 billion, and is expected to rise to $1.435 trillion in 2031[1]. Trucks also play a key role in the multimodal freight transportation network, as critical first-mile and last-mile (FMLM) links[31]. Hence, monitoring and understanding truck activity (including freight information) has become an essential component to effectively support the growth of freight movement.

In this paper, we propose a fine-grained truck classification system as shown in Fig. 1. The system will detect and classify the trucks' body, tractor and trailer based on the Federal Highway Administration's (FHWA) vehicle classification schema[30] shown in Fig. 2. The FHWA schema classifies trucks based mainly on the number of axles and trailers. The input of our system is the raw image of the vehicle on the highway; the image passes through the four stages of the system. In the first stage, the truck will be detected.
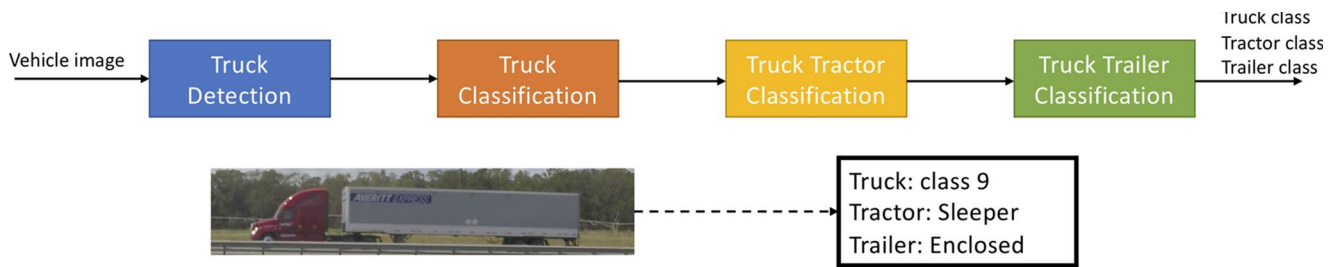
Fig. 1 The truck taxonomy classification pipeline

Then, the remaining three stages will classify the truck's body, the truck's tractor and the truck's trailer accordingly.

The classification is performed using a deep learning technique called Convolutional Neural Network (CNN) [17]. Deep learning is a machine learning technique that has achieved state of the art results in many domains. This includes object detection and classification in images[32], machine translation[27] and natural language processing[16]. This technique consists of a multi-layered artificial neural network that maps patterns into labels and classes. In order to create neural network models that create highly accurate classifiers for our system, a large, labeled image dataset is required to train each model. We gathered and labeled images to train the models for each stage of our system.

In a previous paper, a hybrid deep learning truck and trailer classification approach was used based on their geometric features [10] [9]. The deep learning classification was applied on the main features that distinguish each truck class, which are the number of wheels (a proxy for the number of axles), also the number, aspect ratio and size of the trailers. Afterwards, a decision tree was built on these geometric features to classify the truck class based on the relative spatial relation between these features. In this paper, we investigate the performance of a plain CNN in classifying the truck taxonomy using raw images. Our contributions can be summarized as:

– Achieve an accurate truck taxonomy classification model using only a plain CNN with a sufficient size of an images dataset and transfer learning.
– Create a finer-grained truck taxonomy classification system than previously achieved by incorporating a wider number of truck classes and components (tractors and trailers).
– Use this finer-grained classification system to automatically predict the commodities transported in the trucks. This is achieved by classifying the trailer type and mapping it to the type of commodity that it carries. This is useful for automatically modeling commodity flows in a region.

– Investigate the trucks' salient features learned by the CNN for each class and in each stage by visualizing the decisions of the classifier using the Gradient-weighted Class Activation Mapping (Grad-CAM) method.

To show the accuracy of our system, we show experimental results for each stage of the system on truck images test data.

2 Related Work

Various methods have been used for detecting vehicles on streets including a scanning mask on a video frame[3] and a Spatio-Temporal (ST) map[22]. Some approaches used background extraction techniques[36][38][5][8] and Haar-like features[14]. Recently deep learning techniques were used such as Faster R-CNN[35], selective search on region proposals[2], and Single Shot multi-box detector (SSD) and You Only Look Once (YOLO) system[4].

Vehicle classification has gained wide interest over the last few years. Work was done on classifying the type of vehicle (sedan, SUV, bus, truck, etc.), while other work focused on classifying the make and model of the vehicle. Different types of sensors were used to acquire the inputs, such as laser scanners[34], accelerometers and magnetometers[21], and Weigh-In-Motion (WIM), and Inductive Loop Detectors (ILD)[11]. However, major approaches used cameras for data collection and developed various models including 3D model fitting of a vehicle[33][37] [18], Support Vector Machines (SVMs)[7] [5] [28], and a hybrid dynamic Bayesian network[15]. Deep learning techniques have also been widely used, such as a Recurrent Neural Network (RNN) with a Bag-Of-Visual-Words (BOVW) [12]. However, CNNs were by far the most-used technique for this problem[13][20][6][3]; some used transfer learning with a CNN[38]. Deep learning methods have also been widely used for classifying the model and make of vehicles, such as a Probabilistic Neural Network (PNN)[23] and a CNN[29][35][14]. Previous work used a hybrid approach; a deep CNN was used to classify the geometric features of the truck (the number of wheels, the number, size, and aspect

ratio of the trailers), then a decision tree was used to classify the truck and the trailer based on relative relation between the features [9][10].

The fine-grained truck classification was not widely explored. Images of the trucks were used for the classification based on the FHWA schema using a CNN[2]. Truck and trailer classification was performed using an SSD and YOLO v3[5]. However, previous work had a limited number of the truck taxonomy classes. In our work, we cover a larger number of the track and trailer classes. Also, we classify truck tractors and special types of trucks.

3 Vehicle Images Dataset Preparation

To train accurate classification models for the truck taxonomy, we gathered and labeled a large image dataset of vehicles on highways. The dataset consists of three parts corresponding to the later three stages of the classification pipeline. Statistics of the entire dataset are shown Table 1. The first part contains labeled images of the FHWA truck classes, which are shown in Fig. 2 for the truck classification stage. The second part contains labeled images of truck tractors and special types of trucks for the truck tractor classification stage. The final part contains labeled images of truck trailer categories and subcategories for the truck trailer classification stage. To the best of our knowledge, this is the largest and most detailed image dataset for a truck taxonomy.

The images were gathered with the help of the Florida Department of Transportation (FDOT). A camera was set up on the side of a highway capturing raw videos of the passing vehicles. We chose a side-view vantage point for capturing the videos to have a clear view of the number of axles and trailers of the passing trucks since these features determine the class of the truck. The salient features of the truck tractors and trailers are also present on the side of the truck.

We used YOLO[24], a state-of-the-art object detection system, to extract images of passing vehicles from the raw videos. The system found the bounding box of each passing vehicle in the raw video. Then, the image of the bounding box was extracted. YOLO did not detect objects with an 100% accuracy. Therefore, manual review of all the images was required to remove partially detected objects. Also, the side-view setup of the camera caused problems in detecting vehicles occluded by other vehicles; these images were also removed. This resulted in hundreds of thousands of vehicle images.

Manually labeling this large number of images would be time consuming. To speed up the process, we trained a CNN on a small set of manually labeled images for each stage of the classification pipeline and used it to provide an initial label for the vehicle images dataset. The CNNs used the inceptionl V3 architecture and transfer learning

Table 1 Truck Taxonomy Images Dataset Statistics

(a) FHWA Truck classes dataset

Truck Class	Number of Images
Class 2	22,307
Class 3	15,138
Class 4	270
Class 5	2,036
Class 6	214
Class 7 - 5 axle	39
Class 8 - 4 axle	279
Class 9	14,628
Class 10 - 6 axle	209
Class 11	371
Class 12	310
Class 13 - 7 axle	46
Class 13 - 8 axle	56

(b) Truck tractor dataset

Truck Tractor	Number of Images
Sleeper tractor	6,979
Day cab	2,293
Service truck	589
Pickup truck	12,574
Van	2,689
RV	809
Box truck	307

(c) Truck trailer dataset

Trailer Category	Trailer Subcategory	Number of Images
Flatbed		2,293
Enclosed		1,535
	Conestoga	195
	Drop Frame	3,398
	Side kit	99
	Open top	29
	Tandem enclosed	185
Tank		1,026
	Food grade	74
	Dry bulk	50
Specialty		
	Car hauler	453
	live stock	88
	Dump trailer	56

to improve their accuracy; these methods will be discussed in the next sections. The CNNs achieved an accuracy of around 78%, providing most of the dataset with a correct label. After this step, all the images in the dataset were reviewed, and the labels of the misclassified vehicles were corrected. The same vehicle images were used for each part of the dataset but with a label that corresponds to the class of that part.
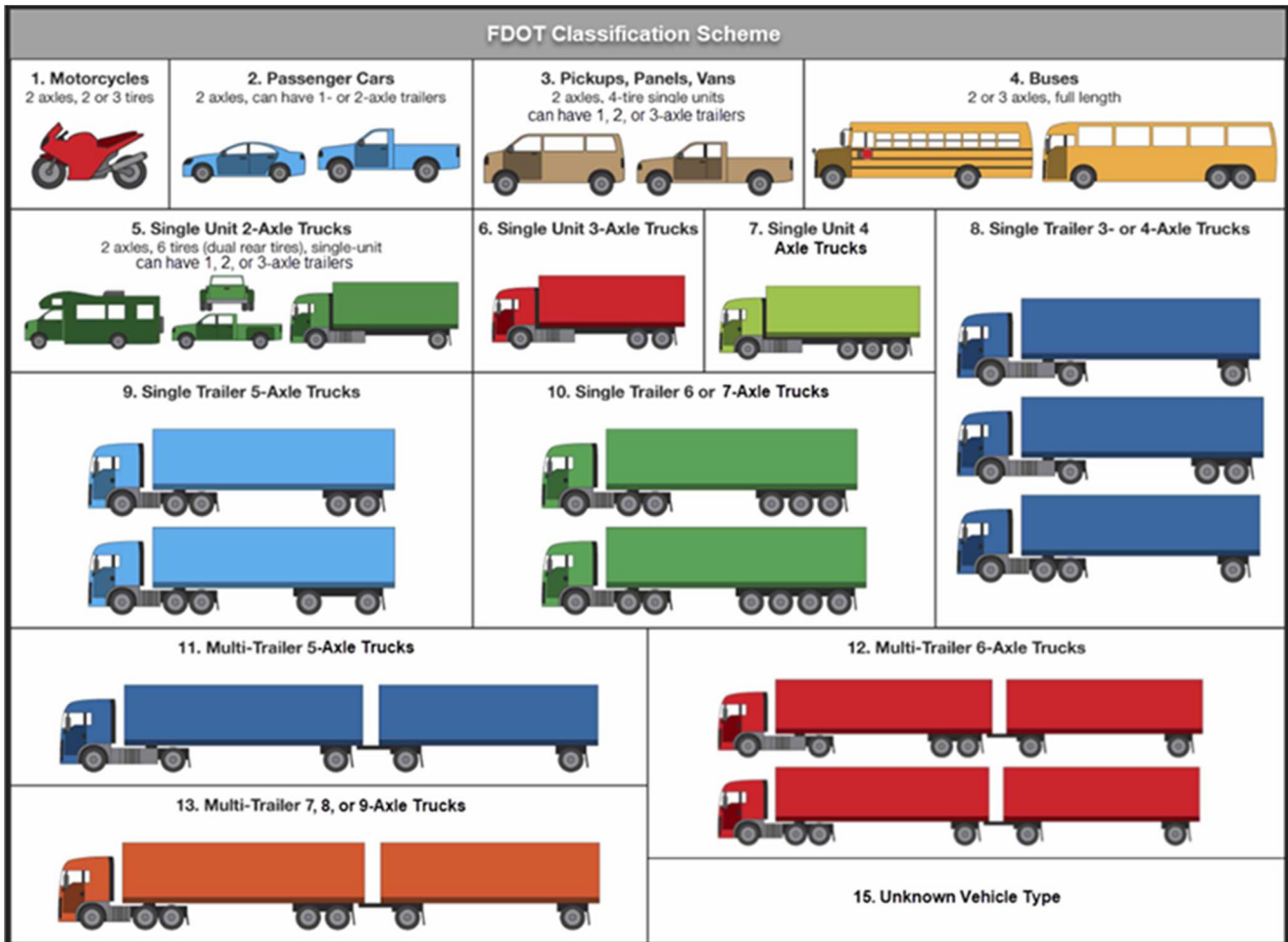
Fig. 2 FHWA Truck Classes

4 Truck Taxonomy Classification System

The system for the truck taxonomy classification consists of four stages. The input to the system is a raw side-view video of the highway. In the truck detection stage, only images of vehicles driving on the highway are extracted from the video. the vehicle image is passed to the truck classification step to recognize the class of the vehicle/truck. In the truck tractor classification stage, the truck tractor or type of truck is determined. Finally, in the truck trailer classification stage, the category or subcategory of the truck's trailer is determined. The first stage uses YOLO object detection while the other stages each use a CNN. These three CNNs were trained on their corresponding part of the dataset. The dataset gathered was largely unbalanced, since some types of truck taxonomy are rarely encountered on highways. Creating a model from an unbalanced dataset will result in a biased, inaccurate model. To alleviate this problem, we used image augmentation techniques on images of trucks from the small classes to increase the number of images in that class. The augmentation techniques generated new images of existing trucks with tilting, flipping, shearing, and other effects applied. Finally, we removed the classes with the very small number of images in them and settled on the rest.

The CNNs used had the Inception V3 architecture, which consist of 48 layers and 24 million parameters. This architecture was chosen for its high accuracy in image classification. However, due to the architecture's high number of parameters, a large training dataset is required to produce an accurate model. Therefore, transfer learning was used with the ImageNet dataset to improve the accuracy and speed up the training time. Only the final two layers were trained on our dataset. This method improves the accuracy of the classifier and greatly reduces CNN training time.

Samples of both the correctly classified and the incorrectly classified truck taxonomy vehicles from our system are shown in Figs. 3 and 4 respectively. The next sections will cover in more detail building the classifier for each stage in the truck taxonomy pipeline.

Fig. 3 Correctly classified truck taxonomy vehicles

4.1 Truck Detection

For truck detection, we use the real-time object detection system YOLO V3. YOLO is a convolutional neural network with 24 convolutional layers and 2 fully connected layers. The system was trained on the CoCo dataset[19] and used to extract vehicle/truck images from the raw video of the highway. For each frame of the video where the system will detect a vehicle, a bounding box on the vehicle will be presented with a class probability score. We will extract the vehicle image from the frame with the highest probability score. The system was not costumized on our dataset since the truck class was already available in the CoCo dataset. Yolo was chosen due to its high-speed in object detection which is necessary for fast moving vehicles on the highway.

4.2 Truck Classification

In the truck classification step, the system should recognize the class of the truck based on the FHWA class scheme. To achieve this task, we used a pretrained inception V3 CNN on the ImageNet dataset and trained only the last two layers on the dataset shown in Table 1a. To balance the data and improve the accuracy of the classifier, classes with a small number of images (e.g., class 7-5 axle, class 13-7 axle, and class 13-8 axle) were removed from the training dataset. Also, classes with low-quality images were removed (e.g. class 8-4 axle and class 10-6 axle). Finally, classes 11 and 12 were combined for similarity, since they are the only classes

with multiple trailers, this resulted in seven truck classes to train our classifier on. Also, undersampling was performed on classes 2, 3, 5 and 9, and oversampling with data augmentation was performed on classes 4, 6 and 11-12 (the multiple-trailers). For each class, about 400 images were chosen for training, about 200 images for validation and 80 images for testing. We trained the CNN for 100 epochs with the RMSProp optimizer with categorical cross entropy as the loss function. The performance of our truck classifier is shown in the experiments section.

4.3 Truck Tractor Classification

In the truck tractor classification step, the tractor type, or the truck's special type is recognized. This step is identical to the truck classification step. A CNN with an Inception V3 architecture was pretrained on the ImageNet dataset. The final two layers only were trained on a subset of the dataset described in Table 1b. To balance the data, undersampling was used on all of the classes except for the Box truck. The size of balanced dataset was 300 images per class. We divided the dataset by choosing 200 images for training, 50 for validation, and 50 for testing. We trained the CNN with the same configuration as in the truck classification step.

4.4 Truck Trailer Classification

In the final step of the system, the truck trailer category and subcategory should be classified from the truck's

Fig. 4 Incorrectly classified truck taxonomy vehicles

image. However, due to the disparity in the frequency of the appearance of many trailers on the highway, we could not collect a sufficient number of images for all the categories and subcategories of trailers that can be used to create a complete trailer classifier. Therefore, we chose only a subset that contained enough images to create an accurate trailer classifier. This subset consists of three main categories of trailers (Flatbed, Enclosed, and Tank trailers) and one subcategory from the specialty trailers (Car hauler).

The CNN for truck trailer classification was created in the same fashion as in the previous two steps. The training dataset consisted of 300 images for each class, and the validation dataset consisted of 100 images for each class. Finally, the model was tested on 50 images for each class.

5 Truck Commodity Prediction

Our truck taxonomy classification system is useful for predicting the commodities contained in trucks, since each trailer type is used for transporting a certain type of commodity. For example, tank trailers are designed to carry liquid goods or gasses, car haulers are used to carry vehicles, enclosed dry trailers are used to carry electronics, or a miscellaneous type of goods. Table 2 shows the mapping between a trailer type and a commodity type. Also knowing the company that the truck is transporting goods for can give a better idea of the commodity contained in the truck, as it narrows the possibilities to the commodities the company distributes and sells. This can be achieved by detecting the company's name and logo on the side of the trailer using a logo detection method.

Table 2 Mapping between trailer category and the commodity type that it carries

Trailer Type	Commodity Type
Car hauler	vehicles
Enclosed	Electronics, misc. commodity type
Tank	Liquid goods or gasses
Flatbed	No commodtity

Predicting the commodities contained in trucks can help find the flow of commodities through a region.

6 Experiments

To evaluate the performance of our truck taxonomy classification system, we will compute the accuracy, precision, recall, and F_1-score for each classification stage of the system. The accuracy is computed on all three datasets (training, validation, and test), while the other metrics are computed only on the test dataset. Also, we will show the confusion matrix for each classification stage to observe the correctly and incorrectly classified truck images for each stage. Top-k accuracy will also be used to measure the accuracy. The most common choice for the value of

k is usually five. However, since the number of classes is small, we chose a smaller number for k. The accuracy and performance results are shown in table 3.

We compare the results of each stage of our system with two machine learning methods, a non-deep learning method, Support Vector Machines (SVMs), and another deep CNN model but with a lower number of layers, VGG-16 model[26]. SVMs were the state-of-the-art in image classification before the advent of deep CNN models. VGG-16 models contain 16 layers in contrast to inception V3's 42 layers, the model won the ILSVR competition in 2014 in image classification of the ImageNet dataset. Also, we compare the results of each stage while using transfer learning for both the VGG-16 and inception V3 architectures and while removing it and training the models with random initializations of the networks' weights.

Table 3 Accuracy results for the truck Taxonomy Classification Stages for our method (inception V3 with transfer learning) compared with SVMs and VGG-16 models, results of the inception V3 and VGG-16 models were evaluated with transfer learning and with random initializations of the networks' weights

Classification Stage	Method	Accuracy	Validation Accuracy	Test Accuracy	Top-k Test Accuracy	Precision	Recall	f1-score
Truck classification	Support Vector Machines	99.42%	-	63.5%	-	62.74%	63.5%	61.14%
	VGG16 without transfer learning	99.94%	90.38	86%	98% (k=3)	86%	86%	86%
	VGG16 with transfer learning	97.85%	78.29%	70%	99.82% (k=3)	73%	70%	71.47%
	Inception V3 without transfer learning	100%	88.86%	41%	58.5% (k=3)	41%	41%	41%
	Inception V3 with transfer learning (**our method**)	100%	92.91%	**89%**	96.24% (k=3)	91%	89%	89%
Truck tractor classification	Support Vector Machines	97.1%	-	51.46%	-	53.65%	51.04%	49.1%
	VGG16 without transfer learning	100%	52.8%	61%	90.1% (k=3)	62%	60%	61%
	VGG16 with transfer learning	95.13%	56%	66%	87.62% (k=3)	65%	64%	64%
	Inception V3 without transfer learning	99.66%	61.07%	70%	93.81% (k=3)	70.85%	69.8%	70.32%
	Inception V3 with transfer learning (**our method**)	99.93%	89.5%	**92%**	99% (k=3)	93%	92%	92.5%
Truck trailer classification	Support Vector Machines	100%	-	72.25%	-	72.23%	72%	72.15%
	VGG16 without transfer learning	100%	72.5%	75.39%	97.91% (k=3)	75.39%	75.39%	75.39%
	VGG16 with transfer learning	100%	94.09%	86.91%	98.95% (k=3)	87.37%	86.91%	87.14%
	Inception V3 without transfer learning	100%	99.55%	90%	98.95% (k=3)	90%	90%	90%
	Inception V3 with transfer learning (**our method**)	100%	99.55%	**96%**	97.91% (k=3)	97%	96%	96%

Our system which uses both the inception V3 model with transfer learning outperforms the other methods in each stage on the test dataset in all metrics. The high top-k test accuracy in our method indicates that the correct prediction is almost always one of the top predictions for all stages. Also, our system achieves a high score on the performance metrics (precision, recall and F_1-score). This shows using deeper CNN models with transfer learning together are essential for obtaining good results in truck taxonomy classification.

Figure 5a-c show the confusion matrix for the test dataset for each classification stage of the truck taxonomy system. The rows correspond to the true labels of each class while the columns correspond to the predicted labels of each class. The number in each cell represents the normalized percentage of images that were predicted as the cell's column class and belonged to the cell's row class. The darkness of the cell represents the magnitude of the percentage. The darkness of the diagonal of the confusion matrix for each stage indicates the high accuracy of our system's prediction. In Fig. 5a, we see that the largest error in our model is in accurately predicting

class 5, which gets confused mostly with class 4. This is attributed to the visual similarity between 2-axle trucks (class 5) and buses (class 4) . Other errors include confusing classes 6 and the joint class 11-12 also due to the visual similarity between single unit trucks and multi-trailer trucks. Finally, our model confuses classes 3 and 4, which is due to the similarity between vans in class 3 with buses in class 4. The highest error in Fig. 5b is in classifying all pickup trucks and mistaking them as boxtrucks or as sleeper tractors, due to the similarity of the front half of those vehicles. Figure 5b shows that our system's trailer classification has some slight trouble in recognizing all flatbeds. This is due to the small side-view body of those trailers and also the existence of different types of cargo on the flatbed might confuse our system.

6.1 Visualizing the Understanding of the Truck Taxonomy Classification System

Since CNNs are black boxes, we cannot interpret the decisions of the classifier in each stage. CNNs can classify

Fig. 5 The confusion matrix for the test dataset for each classification stage of the truck taxonomy system. The value in each cell represents the normalized The darkness of the diagonal in each figure shows the high accuracy of our system across all stages
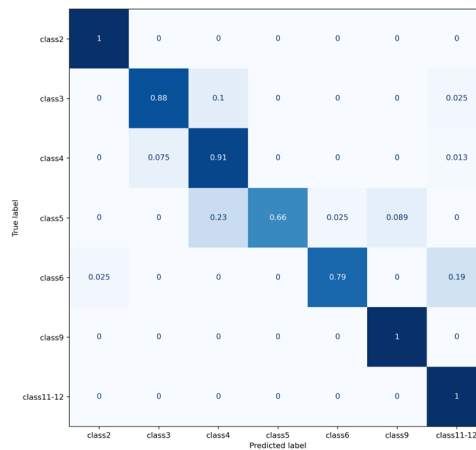
(a) Truck classification
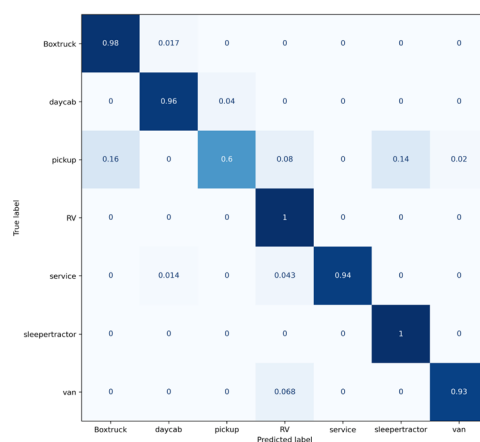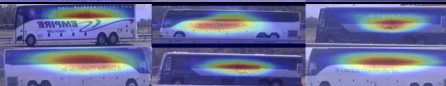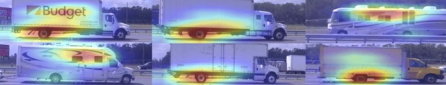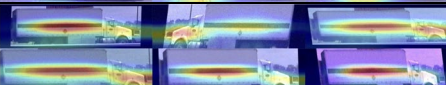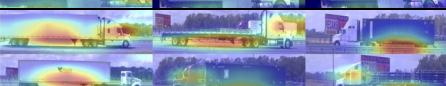
(b) Tractor classification
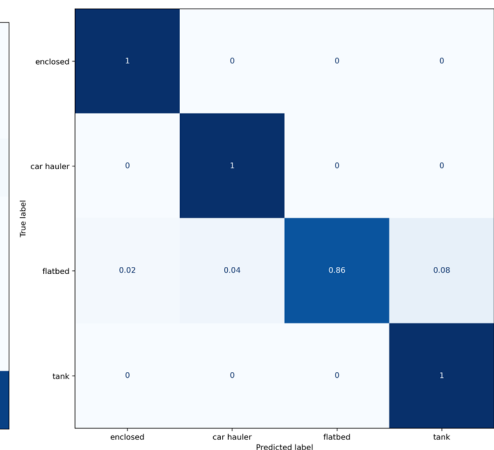
(c) Trailer classification

Table 4 Grad Cam visualizations of samples of the true positives of our truck classification model for each truck class
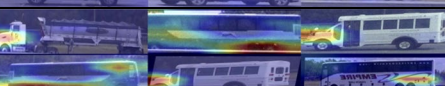

objects correctly, but this classification may be based on non-related features in the image or non-salient features of the object. To investigate what our classifiers have learned from the training and to ensure that the vehicles are being classified based on their salient features, we apply the Gradient-weighted Class Activation Mapping (Grad-CAM)[25] method for our classification stages. The method uses the gradient of a class flowing into the last convolutional layer of the CNN to understand which regions of the image inform its classification decision. Regions of the image that contributed higher to the decision will have higher values for their corresponding pixels. We will visualize these values on the image using a heatmap.

Tables 4-11 show some of the interesting results of applying Grad-CAM on each classification stage for the true positive, false positive and false negative samples. In table 4, we find the truck feature in images for each truck class that

most affects both the correct and incorrect classification decisions. For class 2 (passenger cars), the side windows of the cars are the feature that determines the correct classification. This feature is the main distinguishing feature between passenger cars and trucks, which shows that our model correctly learned the correct feature to determine this class. For class 3 (pickups and vans), the front part of the cargo area for pickup trucks and the wide side panels of vans are the determining features learned, which are mostly considered the distinguishing features for this class. For class 4 (buses), the wide side and wide windows of the buses determine this class, which are the distinguishing feature for this class. For class 5 (2-axle trucks), the single back axle affects the classification the most, this feature is only present in this class. For RVs, which belong to this class, the unique top side of it distinguishes it. Thus, our model learned the unique features

Table 5 Grad Cam visualizations of samples of the false positives of our truck classification model for each truck class

Table 6 Grad Cam visualizations of samples of the false negatives of our truck classification model for each truck class
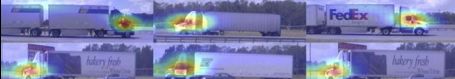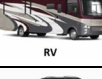
Table 7 Grad Cam visualizations of samples of the true positives of our truck tractor classification model for each tractor class

Table 9 Grad Cam visualizations of samples of the false negatives of our truck tractor classification model for each tractor class

for this class correctly. For class 6 (single unit 3-axle trucks), the whole body of the truck including the axles determines this class. However, this can be confused with the wide body of buses as seen in the false positive samples. For class 9 (single trailer 5-axle trucks), we observe that the space between the front and back axles is the distinguishing feature. However, this can be confused with the same space present in class 5 trucks. For class 11-12 (multiple trailer trucks), the most distinguishing feature is the space between the trailers, which is unique to this class. Therefore, the unique distinguishing feature was correctly learned.

We also want to inspect the features that cause our method to misclassify some of the truck classes by applying Grad-CAM on truck images that either generated a false

Table 8 Grad Cam visualizations of samples of the false positives of our truck tractor classification model for each tractor class

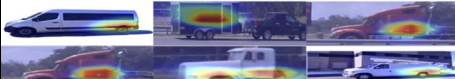**Table 10** Grad Cam visualizations of samples of the true positives of our truck trailer classification model for each trailer class
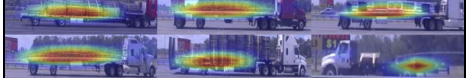
Table 11 Grad Cam visualizations of samples of the false negatives of our truck trailer classification model for each trailer class

Trailer Class	False Negative Samples
Flatbed trailer	

positive and false negative classification on each truck class. Tables 5 and 6 show a sample of false positive and false negative classification for the truck classes respectively. The Features that highly affect the misclassification are features with similar visual appearance among certain truck classes, as also shown in the confusion matrix of the truck classification in Fig. 5a.

Table 7 shows the same results for truck tractors. For day cab tractors, the classification decision was correctly determined by the tractor part of the truck. For sleeper tractors, the wide space of the tractor was the distinguishing feature. For all the true positive samples of the box trucks, the bottom edge of the truck was the determining feature. For the RV class, the wide body of the vehicle was the determining feature. The side window and panel were the distinguishing features of the pickup truck class. In service trucks, the compartments in the bed of the truck were the distinguishing feature. Finally, the whole body of the van was the determining feature. Tables 8 and 9 show the sample of our system's misclassified tractors as false positives and false negatives respectively. They show the visual features that confused our system, which were either similar features in different classes of truck tractors or are caused by the background of the image or occlusion caused by having other vehicles in the same image.

Table 10 shows the Grad CAM results of true positive samples for each trailer classes. For enclosed trailers, the bottom part of the trailer was the determining feature. For flatbed trailers, the whole bed surface was the distinguishing feature. For tank trailers, the tank was correctly determined as the distinguishing feature. For car haulers, the hauler was a determining feature regardless of whether it had cars in it or not. Hence, a robust feature for this class was learned. Table 11 shows the misclassified trailers as false negatives, which were only in the flatbed trailer class. Flatbeds are relatively harder to classify due to their flat shape and having cargo on them or other vehicles will alter their look or cause our system to classify the carried vehicle instead of the trailer.

7 Conclusion

In this paper, we have proposed a truck taxonomy classification system for detecting and classifying trucks, truck tractors, and truck trailers' types. The system uses Yolo for truck detection and deep convolutional neural networks (CNNs) for classification in each stage. A large fine-grained truck image dataset was built to create the system. Images of trucks were gathered and labeled specifying the trucks, tractors and trailers classes. The system can be used to predict the type of commodity the truck is carrying by classifying the truck's trailer and mapping the trailer to the commodity that it usually carries. We have shown that the system achieves a high-accuracy classification in each stage of the truck taxonomy. The classifiers for each stage of the system were analyzed by visualizing their understanding of the learned classes using the Grad Cam method. This led to the discovery of the salient features of each class that informed the decision of the classifiers.

Declarations

References

1. Freight transportation forecast (2020). URL http://www.atabusinesssolutions.com/
2. Adu-Gyamfi, Y.O., Asare, S.K., Sharma, A., Titus, T.: Automated vehicle recognition with deep convolutional neural networks. Transportation Research Record **2645**(1), 113–122 (2017). 10.3141/2645-13. URL https://doi.org/10.3141/2645-13
3. Bautista, C.M., Dy, C.A., Maalac, M.I., Orbe, R.A., Cordel, M.: Convolutional neural network for vehicle detection in low resolution traffic videos. In: 2016 IEEE Region 10 Symposium (TENSYMP), pp. 277–281 (2016). 10.1109/TENCONSpring.2016.7519418
4. Chen, L., Jia, Y., Sun, P.Y., Sinnott, R.O.: Identification and classification of trucks and trailers on the road network through deep learning. BDCAT '19, p. 117–126. Association for Computing Machinery, New York, NY, USA (2019). 10.1145/3365109.3368781. URL https://doi.org/10.1145/3365109.3368781
5. Chen, Z., Ellis, T., Velastin, S.A.: Vehicle detection, tracking and classification in urban traffic. In: 2012 15th International IEEE Conference on Intelligent Transportation Systems, pp. 951–956 (2012). 10.1109/ITSC.2012.6338852
6. Dong, Z., Wu, Y., Pei, M., Jia, Y.: Vehicle type classification using a semisupervised convolutional neural network. IEEE transactions on intelligent transportation systems **16**(4), 2247–2256 (2015)
7. Fu, H., Ma, H., Liu, Y., Lu, D.: A vehicle classification system based on hierarchical multi-svms in crowded traffic scenes. Neurocomputing **211**, 182–190 (2016)
8. Gupte, S., Masoud, O., Martin, R.F.K., Papanikolopoulos, N.P.: Detection and classification of vehicles. IEEE Transactions on

Intelligent Transportation Systems **3**(1), 37–47 (2002). https://doi.org/10.1109/6979.994794

9. He, P., Wu, A., Huang, X., Scott, J., Rangarajan, A., Ranka, S.: Deep learning based geometric features for effective truck selection and classification from highway videos. In: 2019 IEEE Intelligent Transportation Systems Conference (ITSC), pp. 824–830 (2019). 10.1109/ITSC.2019.8917097

10. He, P., Wu, A., Huang, X., Scott, J., Rangarajan, A., Ranka, S.: Truck and trailer classification with deep learning based geometric features. IEEE Transactions on Intelligent Transportation Systems pp. 1–10 (2020). 10.1109/TITS.2020.3009254

11. Hernandez, S.V., Tok, A., Ritchie, S.G.: Integration of weigh-in-motion (wim) and inductive signature data for truck body classification. Transportation Research Part C: Emerging Technologies **68**, 1–21 (2016)

12. Hussain, K.F., Moussa, G.S.: On-road vehicle classification based on random neural network and bag-of-visual words. Probability in the Engineering and Informational Sciences **30**(3), 403–412 (2016). https://doi.org/10.1017/S0269964816000073

13. Huttunen, H., Yancheshmeh, F.S., Ke Chen: Car type recognition with deep neural networks. In: 2016 IEEE Intelligent Vehicles Symposium (IV), pp. 1115–1120 (2016). 10.1109/IVS.2016.7535529

14. Jo, S.Y., Ahn, N., Lee, Y., Kang, S.: Transfer learning-based vehicle classification. In: 2018 International SoC Design Conference (ISOCC), pp. 127–128 (2018). 10.1109/ISOCC.2018.8649802

15. Kafai, M., Bhanu, B.: Dynamic bayesian networks for vehicle classification in video. IEEE Transactions on Industrial Informatics **8**(1), 100–109 (2012). https://doi.org/10.1109/TII.2011.2173203

16. Kamath, U., Liu, J., Whitaker, J.: Deep learning for NLP and speech recognition, vol. 84. Springer (2019)

17. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1, NIPS'12, pp. 1097–1105. Curran Associates Inc., USA (2012). URL http://dl.acm.org/citation.cfm?id=2999134.2999257

18. Lai, A.H., Fung, G.S., Yung, N.H.: Vehicle type classification from visual-based dimension estimation. In: Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE, pp. 201–206. IEEE (2001)

19. Lin, T.Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L., Dollr, P.: Microsoft coco: Common objects in context (2014). URL http://arxiv.org/abs/1405.0312. Cite arxiv:1405.0312Comment: 1) updated annotation pipeline description and figures; 2) added new section describing datasets splits; 3) updated author list

20. Liu, H., Tian, Y., Yang, Y., Pang, L., Huang, T.: Deep relative distance learning: Tell the difference between similar vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2167–2175 (2016)

21. Ma, W., Xing, D., McKee, A., Bajwa, R., Flores, C., Fuller, B., Varaiya, P.: A wireless accelerometer-based automatic vehicle classification prototype system. IEEE Transactions on Intelligent Transportation Systems **15**(1), 104–111 (2014). https://doi.org/10.1109/TITS.2013.2273488

22. Malinovskiy, Y., Wu, Y.J., Wang, Y.: Video-based vehicle detection and tracking using spatiotemporal maps. Transportation

Research Record **2121**(1), 81–89 (2009). 10.3141/2121-09. URL https://doi.org/10.3141/2121-09

23. Psyllos, A., Anagnostopoulos, C.N., Kayafas, E.: Vehicle model recognition from frontal view image measurements. Computer Standards & Interfaces **33**(2), 142–151 (2011)

24. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. arXiv preprint arXiv:1612.08242 (2016)

25. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 618–626 (2017)

26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Y. Bengio, Y. LeCun (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015). URL http://arxiv.org/abs/1409.1556

27. Singh, S.P., Kumar, A., Darbari, H., Singh, L., Rastogi, A., Jain, S.: Machine translation using deep learning: An overview. In: 2017 international conference on computer, communications and electronics (comptelix), pp. 162–167. IEEE (2017)

28. Sochor, J., Herout, A.: Unsupervised processing of vehicle appearance for automatic understanding in traffic surveillance. In: Digital Image Computing: Techniques and Applications (DICTA), 2015 International Conference on, pp. 1–8. IEEE (2015)

29. Sochor, J., Herout, A., Havel, J.: Boxcars: 3d boxes as cnn input for improved fine-grained vehicle recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3006–3015 (2016). 10.1109/CVPR.2016.328

30. Texas DOT: Vehicle Classification Using FHWA 13-Category Scheme, Appendix A:

31. Tok, A., Hyun, K., Hernandez, S., Jeong, K., Sun, Y., Rindt, C., Ritchie, S.G.: Truck activity monitoring system (tams) for freight transportation analysis (2017)

32. Wang, W., Yang, Y., Wang, X., Wang, W., Li, J.: Development of convolutional neural network and its application in image classification: a survey. Optical Engineering **58**(4), 040901 (2019)

33. Xiang, Z., Huang, X., Zou, Y.: An effective and robust multi-view vehicle classification method based on local and structural features. In: Multimedia Big Data (BigMM), 2016 IEEE Second International Conference on, pp. 68–73. IEEE (2016)

34. Xiao, W., Vallet, B., Schindler, K., Paparoditis, N.: Street-side vehicle detection, classification and change detection using mobile laser scanning data. ISPRS Journal of Photogrammetry and Remote Sensing **114**, 166–178 (2016). https://doi.org/10.1016/j.isprsjprs.2016.02.007

35. Yu, S., Wu, Y., Li, W., Song, Z., Zeng, W.: A model for fine-grained vehicle classification based on deep learning. Neurocomputing (2017)

36. Zhang, G., Avery, R., Wang, Y.: Video-based vehicle detection and classification system for real-time traffic data collection using uncalibrated video cameras. Transportation Research Record: Journal of the Transportation Research Board **1993**, 138–147 (2007)

37. Zhang, Z., Tan, T., Huang, K., Wang, Y.: Three-dimensional deformable-model-based localization and recognition of road vehicles. IEEE Transactions on Image Processing **21**(1), 1–13 (2012)

38. Zhou, Y., Nejati, H., Do, T.T., Cheung, N.M., Cheah, L.: Image-based vehicle analysis using deep neural network: A systematic study (2016)