CLASSIFICATION OF TRUCKS USING CAMERA DATA

O MOKONE[1], CC DE SAXE[2]

[1] School of Mechanical, Industrial & Aeronautical Engineering, University of the Witwatersrand, 1 Jan Smuts Ave., Johannesburg, 2000;
CSIR Smart Mobility, 627 Meiring Naudé Road, Pretoria, 0081
Cell: 078 619 5997; Email: wethu@rocketmail.com

[2] CSIR Smart Mobility, 627 Meiring Naudé Road, Pretoria, 0081;
School of Mechanical, Industrial & Aeronautical Engineering, University of the Witwatersrand, 1 Jan Smuts Ave., Johannesburg, 2000
Tel: 012 841 4013; Email: cdsaxe@csir.co.za

ABSTRACT

Understanding the precise movements of different commodities on South African roads can help in not only describing the logistics sector more accurately, but also in the planning of road infrastructure maintenance and investment. Truck combinations can be classified into several classes broadly associated with different commodity groups, including tautliners, tankers, flatbeds (general freight) and flatbed (containerised freight). Current truck classification systems in South Africa are able to classify trucks by number of axles and vehicle mass, but are unable to determine the combination type and hence commodity group. Video data allows for truck combinations to be classified in more detail using image-based classifiers. The latest developments in deep learning algorithms has made it possible for accurate classification of vehicle types using camera data. A CCTV camera feed of a section of the N3 was provided by the South African National Roads Agency Limited (SANRAL), and was used as a case study to develop a proof of concept classifier for tautliner and tanker truck combinations, using a transfer learning approach and the pre-trained ResNet50 classifier. The results indicate good accuracy based on relatively small datasets. Future work will focus on further optimisation and investigating the training dataset requirements in more detail.

Keywords: Heavy vehicles, Trucks, Logistics, Deep learning, Classification, Computer Vision, Transfer Learning, ResNet50

1 INTRODUCTION

1.1 Background

Freight modes in South Africa include road, rail, air, pipelines and coastal. Road is however the most dominant mode, accounting for 75.88% of South Africa’s freight transport in 2013 (Department of Transport, 2017). This can partly be attributed to competitive pricing and flexibility, though policy and legislation have helped influence this as well (Department of Transport, 2017). In 1993, the Department of Transport increased the minimum axial load from 8200 kg to 9000 kg, allowing for more efficient movement of goods by road (Department of Transport, 2017). Even though there is a desire to move freight to rail, it is likely that road freight will remain the dominant mode for the foreseeable future, so implying that much of the country’s logistics movements will be via road. It is important therefore that this freight movement be monitored carefully in order to monitor the state of logistics in the country, and to help authorities plan road infrastructure maintenance and investment.

Detailed knowledge of the types of trucks and associated traffic on the network can help to understand the movement of freight in South Africa, and is useful for economic studies, road planning, maintenance interventions etc. Currently, the South African National Roads Agency SOC Limited (SANRAL) uses a four-class vehicle classification system for toll and infrastructure planning purposes, as summarised in Table 1 (Smith & Visser, 2004).

Table 1: SANRAL vehicle classification

Single loop	Dual loop	Dual loop with axle sensor	Dual loop with sensor (Toll)
None	None	Motorcycle	Toll Class 1
Light	Light	Light motor vehicle	
		Light motor vehicle + trailer	
Heavy	Short Truck	Two axle bus	Toll Class 2
		Two axle single unit	
		Three Axle Unit + trailer (Max axles)	Toll Class 3
		Two axle Single Unit + Trailer (Axles Max)	
		3 Axle Single Unit Including Single Axle Light Trailer	
	Medium Truck	Four or less Axle single trailer	
		Busses with 5 or more axles	Toll Class 4
		Three axle Single Unit and light trailer (more than 4 axles)	
		Five Axle single Trailer	
		Six axle single trailer	
	Long Truck	Five or less axle Multi-trailer	
		Six axle multi-trailer	
		Seven Axle Multi-trailer	
		Eight or more Axle Multi-trailer	

This classification system differentiates trucks on the basis of number of axles and axle loads. This is useful from a toll and road impact point of view, but gives little insight into the types of freight moving on different routes. Therefore the current information has limited value for logistics studies. In recent years, cameras coupled with image-based classification algorithms have seen substantial growth in performance and application, and have been used for a wide range of vehicle-based classification tasks to good effect (Moussa, 2014), including traffic monitoring and accident detection. The technology has the added benefit of requiring little additional equipment (existing CCTV camera feeds can be used), and having very low maintenance requirements compared to, for example, the inductance-based loop detectors currently used (Zhang, et al., 2007).

In this work, a Neural Network based classification system was used. Neural networks are a subset of machine learning algorithms, and in this case the specific method of transfer learning was adopted. These are described in more detail in the following sections.

1.1.1 Machine learning

Machine learning is a subset of the artificial intelligence field, where artificial intelligence describes the use of computers to solve a variety of problems through algorithms of varying complexity. Machine learning is a form of applied statistics that uses computers and algorithms to statistically estimate complex functions (Goodfellow, et al., 2017). Machine learning algorithms can generally be divided into three subsections, namely:

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Supervised learning involves the training of an algorithm using existing labelled data, in order for it to make predictions on new unseen data (Raschka & Mirjalili, 2017). Supervised learning can be further categorised into classification and regression problems. Regression entails making predictions where the outcome is a continuous value, whereas classification involves outputs that are discrete but unordered (Raschka & Mirjalili, 2017).

1.1.2 Transfer learning with Convolutional Neural Networks

Transfer learning is a method of reducing the training requirements of a new classifier, by relying on an existing neural network trained to detect similar classes (Menshawy, 2018). Most of the layers of feature recognition within a neural network identify image features common to most objects, and only the last few layers of the network focus on the object-level identifiers. This means that only the last few layers of the network must be retrained for new classes, while the rest can remain. Transfer learning allows for a significant reduction in the size of the training dataset needed for the neural network to converge (Menshawy, 2018) or reach a stable point for the new classification task. This in turn means reduced training time.

1.1.3 Previous work

A number of studies have shown the potential for vehicle-type classification through image processing, such as the work of (Moussa, 2014), in which geometric and appearance attributes to classify vehicles with the help of support vector machines (SVM). The idea of using small datasets was explored by researchers who modified a deep VGG16 network and trained it on the CIFAR-10 dataset[1]. The dataset had 60 000, 32×32 pixel colour images separated into 10 classes. The performance was not necessarily state-of-the-art but showed the potential of using neural networks to train on relatively small datasets (Liu & Deng, 2015).

“Deep Convolutional Neural Networks (CNN’s) like the VGG16 network allow for improved performance because of the increased number of layers to extract features. A downside of these networks is that they exhibit degradation of the training accuracy as the number of layers is increased which is not caused by data overfitting. Researchers showed that this problem could be reduced with the ResNet family of deep networks (He, et al., 2016). Different configurations of the ResNet network were developed and compared to the VGG network. Amongst these networks were the ResNet 34/50/101/152[2] variations. ResNet 50/101/152 classifiers were considerably more accurate than the ResNet 34 variant.

More recently, transfer learning has been applied to truck classification tasks, using the pre-trained ResNet_152 network (Nezafat, et al., 2018). The network was used as a feature extractor and the following 3 supervised classifiers were compared: K-nearest neighbourhood (KNN), a Support Vector Machine (SVM), and a Multilayer Perceptron (MLP).The network was trained on 1500 images of trucks. The two body types were: an

[1] The dataset is small relative to the ImageNet dataset of 1.2 million images that was used to train AlexNet in 2012 (Krizhevsky, et al., 2012)

[2] These numbers relate to the number of layers in the network

intermodal container truck and a closed body truck shown in Figure 1. The images used were taken from a single camera point of view and had no other cars obstructing the captured trucks. The MLP model achieved an accuracy of 96.5%, with the SVM model coming second with an accuracy of 88% and the KNN model achieving an accuracy of 84,7%.

Figure 1:Sample dataset for truck classification (Nezafat, et al., 2018)

2 OBJECTIVES

The goal of this work was to develop a proof of concept truck classifier for South Africa, based on South African data. The specific objectives were as follows:

1) Develop a truck classifier to distinguish between a fuel tanker and container trucks on South African roads.

2) Assess the performance of the classifier against the following metrics:
   - Amount of training data
   - The effects of image resolution
   - The effects of occlusion and background noise

3 METHODOLOGY

Freeway CCTV video footage was generously provided by SANRAL in a compressed video format. The CCTV footage obtained was for a section of the N3 between Pietermaritzburg and Durban. The video has a resolution of 800×600 pixels and a frame rate of two frames per second. Processing of the data and development of the classifier was carried out in MATLAB, making use of the Deep Learning and Image Processing toolboxes (Mathworks, 2018).

The transfer learning approach in this work used the ResNet50 model that is fine-tuned and used as a classifier and not as a feature extractor (to feed to a supervised classifier like a

SVM). The ResNet50 network was chosen as the base model because of the improvements in accuracy compared with other deep CNN's.

3.1 Pre processing

Raw image data usually require pre-processing in order to yield optimal performance (Raschka & Mirjalili, 2017). Hence, pre-processing was carried out on the video data before processing. Convolutional Neural Networks such as ResNet-50 typically require input images with a 1:1 aspect ratio and ResNet-50 requires input images of resolution 224×224 pixels. Images were cropped to the correct aspect ratio, then if images were smaller or larger in resolution than 224×224, they were scaled accordingly.

For supervised machine learning, training and validation images must be labelled, typically through a manual process. This is the most essential part of supervised learning and special care needs to be taken when the images are labelled to avoid mislabelling. In this work, the images were sorted into folders according to the classes after the pre-processing procedure. The data was then sorted into training, testing and validation sets. "Data hygiene" was emphasised so that the testing data set was only used for testing the classifier and not the training process as well. A breakdown of the image datasets is shown in **Table 2**.

Table 2: Training network data split

Dataset	Percentage	Number of images
Training	60%	177
Validation	20%	29
Testing	20%	29
	Total	294

The pre-trained network was added to the workspace. Once the network was loaded, the size of the first input layer was retrieved in order to further condition the images from the datasets. The first input layer of the pre-trained ResNet50 network has a dimension of 224×224×3. The images were scaled to this resolution.

3.2 Replacing network layers

The next step was to find the layers of the network which will be replaced in the transfer learning task. The final three layers are:

- A "fully-connected" layer (in which every neuron in one layer is connected to every neuron in the next layer)
- A "softmax" layer (a type of "loss" layer, through which numerical inputs are passed through a suitable loss function to create a set of probabilities which sum to 1)
- A classification (or output) layer, which contains the resultant classification result

These layers allow the network to give predictions according to the number of classes provided. The base model has a fully connected layer with a dimension of 1×1×1000 because it was trained to classify 1000 different classes. In this case, the model needs to classify 2 classes.

Table 3 shows the last layers of the ResNet50 pre-trained network. Namely, the fully connected, softmax and classification layers. As described above, the final 3 layers are related to the number of classes to be classified. The modified model will give predictions from 2 classes. Hence, layer 175 will be replaced by a fully connected layer that has a dimension of 1×1×2 followed by the softmax layer with the same dimension.

Table 3: Base ResNet50 network final layers

Layer	Layer name	Layer type	Dimension[3]
175	‘fc1000’	Fully connected	1×1×1000
176	‘fc1000_softmax’	Softmax	1×1×1000
177	‘ClassificationLayer_fc1000’	Classification	

3.3 Training options

The training options are where the “hyper-parameters” are set. Depending on the optimisation algorithm used – i.e. stochastic gradient descent, adaptive moment estimation (“adam”) etc. – the training options vary. For this work, stochastic gradient descent was used, which requires the following options to be set (amongst others):

- How the learning rate changes
- The maximum number of iterations
- The minibatch size
- The validation frequency

The training options used are summarised in **Table 4**.

Table 4:ResNet50 training options

Minibatch size	10
Number of epochs	6
Learning rate	0,0003
Validation frequency	19

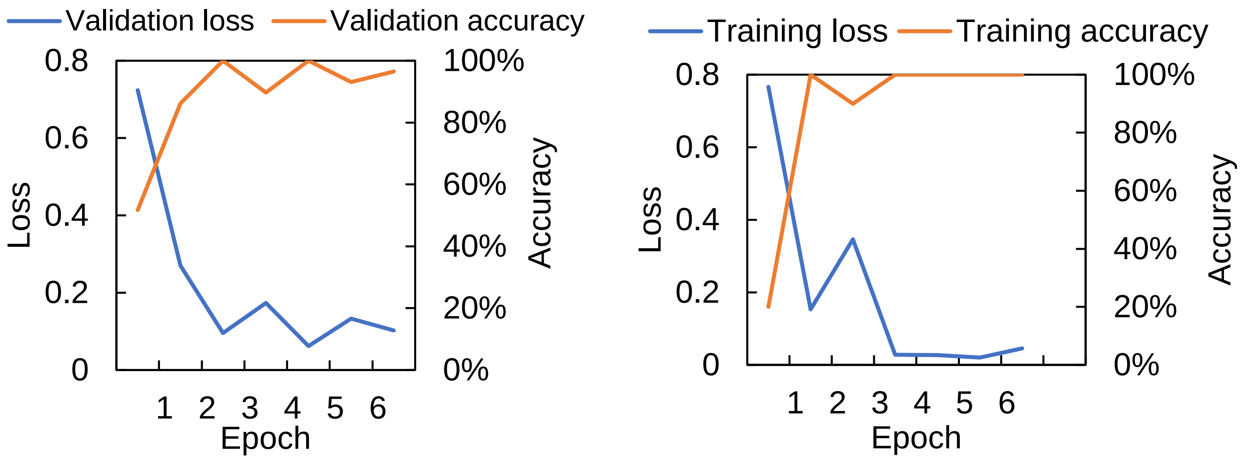
4 RESULTS AND DISCUSSION

4.1.1 Results

After running the training process, the validation and testing accuracy were reported as 96.55% and 98.86% respectively. The accuracy and “loss rate” are useful performance metrics for the classifier. The accuracy represents the percentage of correct predictions in an iteration (or epoch) as the model changes. The cost/loss function (called loss rate) represents the error of the model. The aim is to minimise the loss function and maximise the accuracy of the model through the training iterations.

Figure 2 shows the accuracy and loss rates for both training and validation of the classifier during the training process. It is clear how the model converges to a stable solution after 4 epochs for the training process. The loss and accuracy values for the validation process also reach a solution at the same number of epochs with oscilating values. The training process could be stopped after 4 epochs as a form of an early stopping criteria if the training loss and accuracies are used. In this case the process was run for 6 epochs. The training process ran for 9 minutes and 38 seconds.

[3] These dimensions relate to the size of each layer represented by a 3-dimensional matrix.



(a) (b)

Figure 2: Training process of network, (a) testing dataset and (b) validation dataset

Figure 3 shows selected classified images resulting from the testing, and for each the classifier certainty is also shown. These images were solely from the testing dataset, and so were not "seen" during the training process, thus maintaining data hygiene. All the images were classified correctly. One result however is borderline, the tautliner on the bottom right, which was correctly classified but with an accuracy of 50.2%. This means that the algorithm gave a 49.8% probability of the image being a tanker truck. This may be due to the resolution of this image, as it had a low initial resolution, which was scaled up for training. Additionally, there is evidence of some occlusion from a car in front of the truck, which may also have influenced the result.

Figure 3: 16 image classifications from the retrained ResNet-50 classifier

Overall, an accuracy of 98.86% was produced from the test dataset. This is a very high accuracy, though this is only a very small dataset. Though the result is promising as a proof of concept. The testing accuracy of 98.86% compares well previous work. Previous work on truck classification achieved a 96.5% accuracy rate (Nezafat, et al., 2018). The higher accuracy achieved here could be attributed to the lower number of images in the dataset. Having a small sample size means that there are less images that test the models limitations. Another factor could be the specific classes of truck being classified. The classes in this paper have distinct appearances compared to the trucks that were studied in the work of Nezafat et al. The model shows that for the given camera angle and training data, it can classify the trucks reasonably accurately from the small training data.

Figure 4 shows an example of a misclassified image. The first image was misclassified as a tanker. This can possibly be attributed to its relatively low resolution (100x100 compared to 271 x 271 for the image on the right) or to how the cropped image has excluded parts of the rear trailer.

Figure 4: Classified images with misclassification

5 CONCLUSIONS AND RECOMMENDATIONS

5.1 Conclusions

1) The modified ResNet50 network, retrained on CCTV image data from SANRAL, demonstrated an accuracy of 98.86%. This is a promising result for a proof of concept of a general camera-based truck classification system for South Africa.
2) The work demonstrates the advantages of transfer learning when training data is limited. In this case a relatively small dataset of 294 images was used to train a two-class truck classifier using a pre-trained ResNet50 network.
    a) Effects of image resolution were noteworthy. It is better to downscale an image to the input size of the network as compared to upscaling it. Upscaling results in reduced effective resolution, which negatively impacts the classification accuracy.
    b) Occlusion and background noise had a relatively small effect on the performance of the classification. This could be because the underlying pre-trained ResNet-50 network has already been trained on a wide variety of scenarios, including those with occlusion.

5.2 Recommendations

The performance of the network on small datasets needs to be further investigated. The network performs well on this small set but the robustness of it can be improved. More training data would be beneficial, but an investigation into precisely how much training data is required would be valuable.

The next step is to train the classifier on additional truck classes, such as flatbeds, car-carrier, side-tippers etc. Additional training data may be required, potentially from different road sections.

The effect of camera position and orientation relative to the traffic requires further investigation. It would be better if a classifier was trained on a variety of views, such that a single classifier could be used by all traffic cameras regardless of location and orientation.

Further work on making the identifications faster could be looked at. Detectors such as the Region Convolutional Neural Networks (R-CNN) and the family of fast and faster R-CNN could be looked at for their applicability.

6 ACKNOWLEDGEMENTS

The authors would like to thank SANRAL for providing the CCTV video footage for use in this work. This work was partially funded by the Department for Science and Technology, though the CSIR’s parliamentary grant funding programme.

7 REFERENCES

Department of Transport, 2017. *Documents.* [Online]
Available at:
https://www.transport.gov.za/documents/11623/39906/7_FreightTransport2017.pdf/a3f7cb55-8d77-4eea-b665-4c896c95a0d8
[Accessed 2 September 2019].

Goodfellow, I., Bengio, Y. & Courville, A., 2017. *Deep learning.* Massachusetts: MIT Press.

He, K., Zhang, X., Ren, S. & Sun, J., 2016. *Deep Residual Learning for Image Recognition.* Las Vegas, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

Krizhevsky, A., Sutskever, I. & Hinton, G. E., 2012. ImageNet Classification with Deep Convolutional. *Advances in neural information processing systems,* p. 1097–1105.

Liu, S. & Deng, W., 2015. *Very Deep Convolutional Neural Network Based Image Classification Using Small Training Sample Size.* Kuala Lumpur, IEEE Computational Intelligence Society.

Mathworks, 2018. *Matlab R2018b.* Natick, Massachusetts: Mathworks.

Menshawy, A., 2018. *Deep learning by example.* Birmingham - Mumbai: Packt.

Moussa, G. S., 2014. Vehicle Type Classification with Geometric and Appearance Attributes. *International Journal of Architectural and Environmental Engineering,* 8(3), pp. 278-282.

Nezafat, R. V., Salahshour, B. & Cetin, M., 2018. *Classification of truck body types using a deep transfer learning approach.* Maui, Institute of Electrical and Electronics Engineers.

Raschka, S. & Mirjalili, V., 2017. *Pyhton Machine Learning; Machine learning and Deep learning with Python, scikit-learn and TensorFlow.* Birmingham: Packt Publishing.

Smith, A. & Visser, A., 2004. *A South African Road Network Classification Based On Traffic Loading.* Johannesburg, Document Transformation Technologies .

Zhang, G., Avery, R. P. & Wang, Y., 2007. Video-Based Vehicle Detection and Classification System for Real-Time Traffic Data Collection Using Uncalibrated Video Cameras. *Journal of the Transportation Board,* 1993(1), pp. 138-147.