URBAN TRAFFIC FLOW ANALYSIS BASED ON DEEP LEARNING CAR DETECTION FROM CCTV IMAGE SERIES

M. V. Peppa [1]*, D. Bell [1], T. Komar[1], W. Xiao [1]

[1] School of Engineering, Newcastle University, Newcastle upon Tyne, UK -
(maria-valasia.peppa, d.bell5, tom.komar, wen.xiao)@ncl.ac.uk

Commission IV, WG IV/6

KEY WORDS: Computer vision, machine learning, tensorflow, traffic monitoring, CCTV big data, infrastructure management

ABSTRACT:

Traffic flow analysis is fundamental for urban planning and management of road traffic infrastructure. Automatic number plate recognition (ANPR) systems are conventional methods for vehicle detection and travel times estimation. However, such systems are specifically focused on car plates, providing a limited extent of road users. The advance of open-source deep learning convolutional neural networks (CNN) in combination with freely-available closed-circuit television (CCTV) datasets have offered the opportunities for detection and classification of various road users. The research, presented here, aims to analyse traffic flow patterns through fine-tuning pre-trained CNN models on domain-specific low quality imagery, as captured in various weather conditions and seasons of the year 2018. Such imagery is collected from the North East Combined Authority (NECA) Travel and Transport Data, Newcastle upon Tyne, UK. Results show that the fine-tuned MobileNet model with 98.2% precision, 58.5% recall and 73.4% harmonic mean could potentially be used for a real time traffic monitoring application with big data, due to its fast performance. Compared to MobileNet, the fine-tuned Faster region proposal R-CNN model, providing a better harmonic mean (80.4%), recall (68.8%) and more accurate estimations of car units, could be used for traffic analysis applications that demand higher accuracy than speed. This research ultimately exploits machine learning alogrithms for a wider understanding of traffic congestion and disruption under social events and extreme weather conditions.

1. INTRODUCTION

1.1 Background

Traffic monitoring and analysis are crucial for an efficient urban planning and management of road traffic infrastructure. An ideal monitoring system would measure traffic movement, density and interactions as well as predict disruption and congestion in order to mitigate hazards for a safe functional infrastructure.

Traditionally, inductive loops embedded in road’s surface and automatic number plate recognition (ANPR) systems are utilised for vehicle detection and travel times estimation (Li, 2008). However, ANPR systems do not always offer an overall extent of road users due to specific focus on recognising characters in car plates and discarding candidate detections if car plate is not fully recognised (Buch et al., 2011). Such limitations have been overcome with the use of state-of-the-art open-source deep learning technologies (Shi et al., 2017) alongside the freely-available closed-circuit television (CCTV) datasets, enabling training of models for detection and classification of various road users as well as traffic monitoring and prediction (Lv et al., 2015). Big data analytics from CCTV image series can provide substantial knowledge of urban traffic flow behaviour. Number of cars, constituting a main parameter in traffic analysis, can be estimated at various time scales from spatially heterogeneous CCTV locations. This type of multi-scale spatiotemporal observations supports understanding of traffic congestion before, during and after a disruptive event.

1.2 Related work

Recent studies have adopted various deep learning neural network systems for traffic analysis. For example Tang et al. (2017) adapted a single shot multibox detector (SSD) to extract the location and orientation of cars from aerial imagery. Wang et al. (2017) used a region proposal convolutional neural network (R-CNN) to classify vehicle types from images of camera sensors which were set up at crossroads. These neural network systems consist of multiple neurons with learnable filters, which are activated after processing an input image using various convolutional operations (e.g. gradients, blobs, edge filters etc.) in combination with fundamental learnable variables (weights and biases). The trained network of neurons can localise specific types of features at any spatial position of the input image.

Today’s architecture of convolutional neural networks (CNNs) was firstly introduced in 1990s by LeCun et al. (1989) to recognise handwritten numbers. This architecture was improved more recently by Krizhevsky et al. (2012), using large benchmark image datasets collected from the web. Further improvements led to state-of-the-art CNNs of better performance with the introduction of region proposals creating, among others, for example the faster region based convolutional neural network (Faster R-CNN; Ren et al. (2017)). R-CNN segments each convolutional layer using a sliding window into proposed regions which contain the predicted location of multiple anchor boxes of different sizes and scales per object. A global optimisation refines the predicted box location with the aid of regression algorithms. In comparison, SSD, introduced

* Corresponding author

https://doi.org/10.5194/isprs-archives-XLII-4-499-2018 |

by Liu et al. (2016), utilises anchor boxes to locate a feature by dividing an input image into grids and using multiple convolution layers of various scales and aspect ratios (Liu et al., 2016; Tang et al., 2017). A mobile architecture of SSD, named MobileNet (Howard et al., 2017), was specifically designed for mobile processing in real time applications.

Various open-source benchmark image datasets with manually labelled features have contributed to the advance of the aforementioned deep learning neural network systems (Gauen et al., 2017). Two of such benchmarks were used here, namely a) Common Objects in Context (COCO; Lin et al. (2014)) and b) Karlsruhe Institute of Technology and Toyota Technological Institute at Chicago Object Detections (KITTI; Geiger et al. (2012)). COCO is a dataset introduced by Microsoft and consists of 80 common objects in a 2.5 million labelled images from the natural environment (Lin et al., 2014; Gauen et al., 2017). KITTI contains datasets from urban environment captured from a car driven around the city of Karlsruhe (Geiger et al., 2012). Over the last few years, these datasets have been used to train the aforementioned deep learning neural network systems with the scope to establish open-to-the-public models from out-of-the-box frozen graph inferences (Rathod et al., 2018). These can ultimately expedite and facilitate object detection and classification for various applications.

1.3 Aim

The study aims to analyse traffic flow pattern using long time CCTV image series. It focuses on testing pre-trained state-of-the-art deep learning neural network systems for car detection from low quality images captured in various environmental conditions. This analysis ultimately exploits machine learning alogrithms for a wider understanding of the traffic behaviour variations under social events and extreme weather conditions.

2. METHODOLOGY

The study presented here investigates the implementation of deep learning methods for car detection in CCTV datasets of the North East Combined Authority (NECA) Travel and Transport Data (NECA, 2018b), Newcastle upon Tyne, UK. Traffic counts are analysed on imagery from two sets of dates and two key locations that connect the city centre with southern and northern suburbs. First analysis was performed at A167 High Street, Tyne Bridge, Gateshead (NECA, 2018c) on Saturday 21st and Saturday 28th of April 2018. The latter is the date of a football match event. Second analysis was performed on 26-28th of February 2018, that is before and during a snow event, taken on A193 Newbridge Street Roundabout where a CCTV sensor monitors the traffic over a part of the A167 Central Motorway (NECA, 2018a). Results from February were compared against traffic patterns analysed on 09-11th of July 2018 with sunny and dry weather conditions, retrieved from the same CCTV sensor location. Traffic flow was finally estimated on a daily-basis on 09-17th of July 2018 from the same CCTV sensor location in order to compare the traffic pattern between weekend and working days.

Experiments were undertaken on two types of supervised neural network systems, namely the SSD and the Faster R-CNN, as adopted in TensorFlow deep learning framework (Shi et al., 2017). Two sets of experiments were conducted, as follows: a) the first set included seven tests (Table 2) to optimise relevant parameters such as learning rate and number of images used for fine-tuning SSD MobileNet V1 COCO (Rathod et al., 2018); b) the second set included three tests (Table 3 and 4) to evaluate the performance of SSD MobileNet V1 COCO, Faster R-CNN COCO and KITTI pre-trained models as fine-tuned on domain-specific imagery. Such local imagery used to aid optimising the performance of deep learning classifiers. It should be noted that various settings required to configure the deep learning process (e.g. image resolutions, intersection over union (IoU) threshold, loss function, number of convolution layers, number of region proposals, filter shapes and sizes, biases, weights etc.) were kept as default, based on the configurations of pre-trained models (Rathod et al., 2018). Finally, traffic flow behaviour was analysed using the fine-tuned models which provided the fastest and most precise performance.

2.1 First experimental set

In particular, in the first experimental set of tests to train the deep learning classifiers and validate classification results, 550 images were sampled from NECA on different days, varying times, different weather conditions and various CCTV sensors around the city, including datasets with noise. Such noise is typically caused by sensor moving/swaying, direct sun rays, reflected sun rays, image blur/saturation and material on the lens (Buch et al., 2011). Hence, this selection would reduce a likelihood of bias in neural network training. 500 images were used for training and 50 images for evaluation. 2140 cars were manually identified, labelled with bounding boxes and organised in appropriate formats for the training process. 384 cars detected in the 50 images, served as ground truth observations for assessing post training model performance. The tests in the first experiment were undertaken with 20,000 training epochs. 500 images were used for all tests in Table 2 with the exception of test 6 and 7. 250 images were used for test 6, whereas no fine-tuning performed for test 7.

2.2 Second experimental set

In the second experimental set two comparisons between the SSD MobileNet V1 COCO, Faster R-CNN COCO and KITTI fine-tuned models were performed. Firstly, a comparison after fine-tuning with 20,000 and 40,000 training epochs and secondly a comparison after fine-tuning with 500 images and then with 700 images. These tests aimed to evaluate the post-training model performance after longer training with a larger training local image dataset. The 500 images were identical to those used for training in the first experimental set. Additional 200 images were sampled from NECA and Urban Observatory (UO; UO (2018)). These included images with views of roads outside the city centre and from dates with higher rainfall than the aforementioned 500 images. Roads of different speed categorisation were also considered (e.g. motorways, regional roads etc.). To identify rainy days, observations from UO were obtained (James et al., 2014). Weather conditions of the selected dates for the 700 images are reported in Table 1.

No of images	Weather conditions	Time of day	Date
30	Cloudy without rain	Morning, midday	28/01/2018
138	Very snowy	Early afternoon	01/03/2018
137	Snowy, dark	Late evening	03/03/2018
137	Snowy, overcast	Midday	04/03/2018
138	Dry, sunny	Late afternoon	15/03/2018 17/03/2018
120	Overcast, rainy	Morning-Night	16/03/2018

Table 1. 700 images extracted from NECA and UO, used for training the deep learning classifiers.

A variety of weather conditions were selected mainly during March 2018. Hence, the deep learning classifiers were trained with domain-specific imagery of spectral heterogeneity. This additional dataset also resulted in a better spatial distribution of CCTV sensors around the extended area of Newcastle upon Tyne. In total 3676 cars were manually identified and labelled from the 700 images.

Moreover, 50 images, different from those used in the first experimental set, were selected for evaluation, with 381 cars identified. These images were obtained from NECA and UO databases at various times during rainy and snowy days, also different from those of the training dataset used in the first experimental set. Specifically, the selected images were captured on 27/02/2018, 02/04/2018, 18/04/2018 and 02/06/2018. This ground truth dataset also included cases with cropped images showing parts of the field of view from CCTV sensors. This is a common issue possibly due to electronics faults of the CCTV sensor, usually occurring during extreme weather conditions.

Initial pre-trained neural network checkpoints were retrieved from Rathod et al. (2018) and used for training at all experiments. After training, fine-tuned neural network checkpoints at 20000 and 40000 epochs were extracted and utilised for performance evaluation with the aid of ground truth of the 50 aforementioned images. For this evaluation three indices were calculated, namely: a) precision which indicates the percentage of identified cars (i.e. true positives) out of the total number of all identified features (i.e. true and false positives); b) recall which indicates the percentage of identified cars out of the total number of cars manually identified (i.e. true positives and false negatives); and c) harmonic mean (F) which is calculated based on Equation 1, as seen below:

$$F = 2\frac{precision \cdot recall}{precision + recall} \tag{1}$$

3. RESULTS

3.1 First experimental set

Results derived from the first experimental set are summarised in Table 2. In terms of the learning rate parameter, which regulates the training process, values lower than 0.01 provided better performance. Precision and recall degraded when half of the images were used for training, as evidenced in test 6 (Table 2).

Test	Learning rate parameter	Identified features	True positives	False positives	False negatives	Precision (%)	Recall (%)	F (%)
1	0.0001	256	246	10	138	96.1	64.1	76.9
2	0.001	252	242	10	142	96.0	63.0	76.1
3	0.01	258	237	21	147	91.9	61.7	73.8
4	0.1	0	0	0	384	0	0	0
5	0.004	210	208	2	176	99.0	54.2	70.0
6	0.0001	242	228	14	156	94.2	59.4	72.8
7	-	183	143	40	241	78.1	37.2	50.4

Table 2. Post training evaluation of the first experimental set with fine-tuned SSD MobileNet V1 COCO.

In test 7, as no training with NECA images was performed, the initial pre-trained model suffered from both poor precision and recall. This indicates the importance of further training already pre-trained systems with images from a local urban environment as it enhances the neural network performance. Among all tests test 1 provided the best indices with a greater number of true positives and a smaller number of false negatives. Based on these findings, a learning parameter of 0.0001 was set up for all remaining tests in the presented research.

3.2 Second experimental set

The second experimental set was performed on a computing platform with characteristics as follows: NVIDIA Quadro P5000, CPU E3 – 1535M v6 @3.10GHz with 64GB RAM. The hours required for training are recorded in Table 3. The description of the neural network pre-trained models are also presented in Table 3. It should be noted that no other process was running while training with the exception of evaluation and visualisation on TensorBoard (TensorBoard, 2018). TensorBoard constitutes a platform to monitor the training of deep learning classifier with the aid of graphs. It also shows the detectors performance on selected images during training. These processes might have consumed part of the computing memory usage as they were performed simultaneously alongside training.

Model	Description	No of images /epochs [K]	Training [hours]	Post training evaluation [seconds/image]
A	SSD MobileNet V1 COCO	500/20	8	1.4
B	Faster R-CNN ResNet 101 COCO	500/20	3.7	7.3
C	Faster R-CNN ResNet 101 KITTI	500/20	3	7.2
A	SSD MobileNet V1 COCO	500/40	15.7	1.4
B	Faster R-CNN ResNet 101 COCO	500/40	10.7	7.2
C	Faster R-CNN ResNet 101 KITTI	500/40	12.2	7.2
A	SSD MobileNet V1 COCO	700/20	8	1.4
B	Faster R-CNN ResNet 101 COCO	700/20	4	7.3
C	Faster R-CNN ResNet 101 KITTI	700/20	6	7.3
A	SSD MobileNet V1 COCO	700/40	15.7	1.4
B	Faster R-CNN ResNet 101 COCO	700/40	6.5	7.2
C	Faster R-CNN ResNet 101 KITTI	700/40	12.2	7.3

Table 3.Details of training time [hours] and evaluation time [seconds] with the 50 images of the second experimental set.

As seen in Table 3 Faster R-CNN ResNet models were trained much faster than SSD MobileNet especially for 20000 epochs. For these epochs the larger image dataset (700) did not significantly affected the training duration, apart from Faster R-CNN ResNet 101 KITTI. This model required double the time when training with 700 images. However, double training epochs resulted in double time for deep learning with the exception of Faster R-CNN ResNet 101 COCO pre-trained model. Surprisingly, this model required only 6 and a half hours for training with 700 images at 40000 epochs whereas the other models needed almost double the time.

Table 3 also reports the seconds per image required for car detection during evaluation of each fine-tuned model. It should be noted that post-training evaluation was performed with CPU only. The detection was processed for every single image out of the 50 images, used as ground truth. As opposed to training duration, SSD MobileNet provided the fastest detections. Whereas, Faster R-CNN ResNet required approximately five times longer period than SSD MobileNet for car detection. It is noteworthy that no significant change on evaluation duration occurred when COCO or KITTI datasets were used as seen in Table 3.

Post training evaluation indices of the second experimental set are presented in Table 4. In particular, models A fine-tuned from SSD MobileNet V1 COCO provided similar performance regardless of the number of images and epochs used for training. Surprisingly, no improvement in precision, recall and harmonic mean was observed when training performed with more images and for longer period (700/40, test 4 in Table 4). Among four tests with model A, SSD MobileNet (500/40, test 2 in Table 4) detected more true positives, resulting in the highest precision, with a relatively low number of false positives, providing the best recall.

Model	Test	No of images /epochs [K]	Identified features	True positives	False positives	False negatives	Precision (%)	Recall (%)	F (%)
A	1	500/20	226	219	7	162	96.9	57.5	72.2
B		500/20	271	254	17	127	93.7	66.7	77.9
C		500/20	253	247	6	134	97.6	64.8	77.9
A	2	500/40	227	223	4	158	98.2	58.5	73.4
B		500/40	242	231	11	150	95.5	60.6	74.2
C		500/40	278	236	42	144	84.9	62.1	71.7
A	3	700/20	222	199	3	131	89.6	52.0	66.0
B		700/20	266	258	8	123	97.0	67.7	79.8
C		700/20	277	246	31	135	88.8	64.6	74.8
A	4	700/40	244	213	6	117	87.3	55.9	68.2
B		700/40	294	277	17	104	94.2	72.7	82.1
C		700/40	271	262	9	119	96.7	68.8	80.4

Table 4. Post training evaluation of the second experimental set.

Models B fine-tuned from Faster R-CNN ResNet 101 COCO show comparable performance. Precision and recall were estimated at four tests always over 93% and 60% respectively. A significant increase in number of true positives and decrease in number of false negatives was observed when longer duration training with more images was performed. However, results of models C fine-tuned from Faster R-CNN ResNet 101 KITTI did not show such a progressive pattern. Evaluation indices from test 1 (Model C 500/20) and test 4 (Model C 700/40) provided comparable values. Among all results from models C, test 4 showed the highest harmonic mean indicating a good balance between precision and recall.

Evaluation indices in Table 3 showed that Faster R-CNN (Models B and C) gave a generally better performance than SSD MobileNet (Models A). This is also evidenced in Figure 1 as more cars were identified with Faster R-CNN (Figure 1b and 1c versus 1a) on images when compared against ground truth (Figure 1d). This image is partially blurred due to stained CCTV sensor during rainfall. It also shows only distant objects as part of the image was not recorded (hence the grey colour in the foreground). SSD MobileNet failed to identify small cars. Moreover, it did not localise the cars’ boxes as precisely as Faster R-CNN. Between the two models fine-tuned with Faster R-CNN (Figure 1b and 1c), Faster R-CNN ResNet 101 COCO provided more accurate detection as it identified a car which is blurred and partly viewed in this particular image.

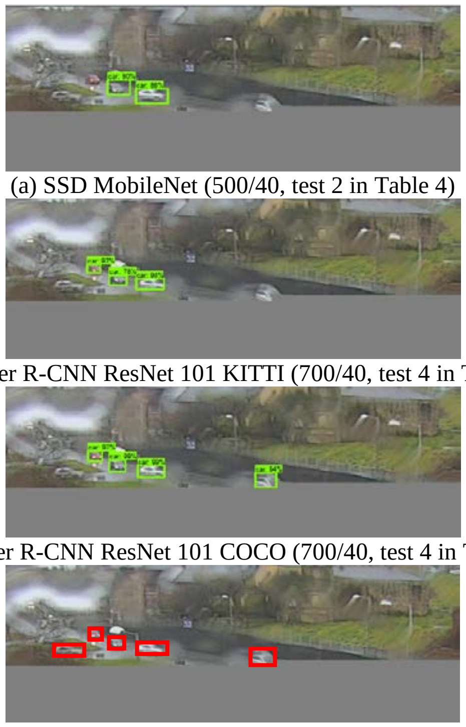

(a) SSD MobileNet (500/40, test 2 in Table 4)
(b) Faster R-CNN ResNet 101 KITTI (700/40, test 4 in Table 4)
(c) Faster R-CNN ResNet 101 COCO (700/40, test 4 in Table 4)
(d) Ground truth

Figure 1. Car detection example on cropped image with blur.

Figure 2 depicts an example of an image captured at night during rainfall resulting in suboptimal lightning conditions and noise due to reflections. SSD MobileNet (Model A, test 2 500/40 in Table 4) failed to detect the one and only car, whereas both Models B and C (test 4 700/40 in Table 4) succeeded (Figure 2). The poor performance of SSD MobileNet showed in the two examples, can be attributed to low complexity (“shallowness”) of SSD as well as to low exposure in such examples with noise on imagery. However, it should be noted that among all tests with model A only SSD MobileNet from test 4 in Table 4 successfully detected this car. This is because more images of rainy days were added into the initial training image dataset. (Section 2.2)

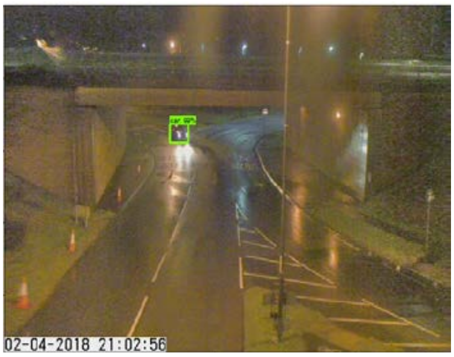
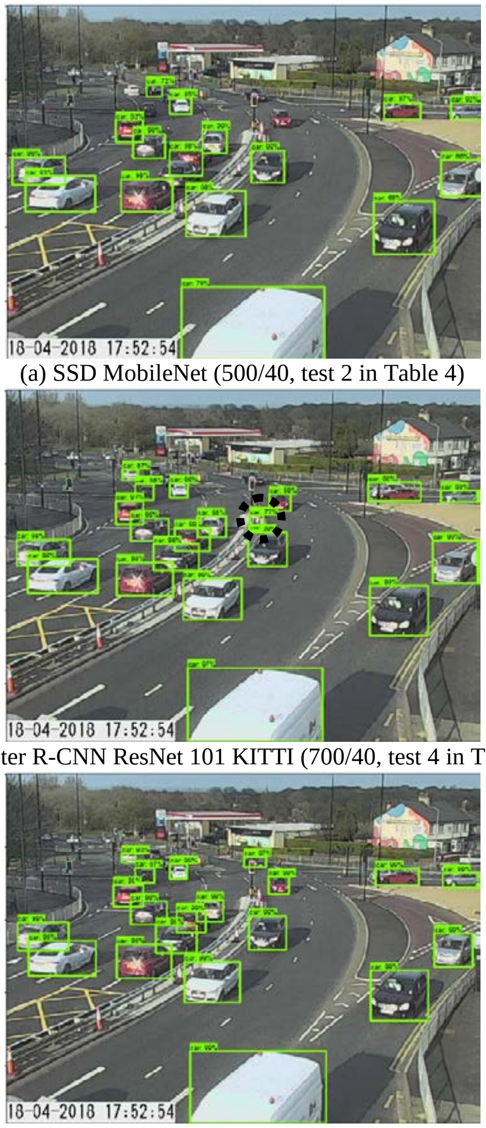


Figure 2. Car detection example during night on a rainy day with Faster R-CNN ResNet 101 COCO and KITTI (700/40, test 4 in Table 4).

A third example of an image with good lighting conditions during a dry sunny day is shown in Figure 3. Here, it was expected that zero false positives and negatives would be calculated. However, Faster R-CNN ResNet 101 KITTI identified a cone nearby the road as a car, (highlighted with a black dotted circle in Figure 3b). Similar errors (e.g. double detection per single car, sensor stain drops detected as car etc.) were observed in the results for all models fine-tuned with Faster R-CNN ResNet 101 KITTI (Table 4), regardless of the training epochs and number of images. This is also reflected with the higher value of false positives in Table 4 compared to those values from all remaining tests.



(a) SSD MobileNet (500/40, test 2 in Table 4)

(b) Faster R-CNN ResNet 101 KITTI (700/40, test 4 in Table 4)

(c) Faster R-CNN ResNet 101 COCO (700/40, test 4 in Table 4)

Figure 3. Car detection example on a dry sunny day.

In Figure 3a zero false positives were observed with SSD MobileNet, but five false negatives were identified. SSD MobileNet failed to detect small cars in the image background, as also seen in previous examples. This type of error was not eliminated even after training with more images and/or epochs.

3.3 Traffic analysis

Based on the findings of the previous two experimental sets, traffic analysis was performed from two fine-tuned models, namely a) SSD MobileNet V1 COCO (Model A, test 3, 500/40 in Table 4) and b) Faster R-CNN ResNet 101 COCO (Model C, test 4, 700/40 in Table 4). The former gave the best performance among all fastest models (see post training evaluation in Table 3 and Table 4). The latter provided the highest precision and recall among all models in Table 4. Faster R-CNN ResNet 101 KITTI was excluded from the following examples, as it did not provide consistent performance in all tests in Table 4.

Figure 4 shows traffic behaviour on Saturday, day of a football match event (28.04.18) and on the previous Saturday (21.04.18) when no social event was organised in the city. During the morning (from 07.00 till 12.00) of the date of the football match the number of cars was gradually increased (as seen in blue in Figure 4). It should be noted that the football match commenced at 15.00 on 28.04.18. Whereas in the morning of the previous Saturday, the number of cars significantly varied (as seen in red in Figure 4). That day a general sinusoidal pattern indicated cars driving in and out of the city centre at certain times. This pattern was not observed on 28.04.18 as it is likely that more cars were parked in the city centre to watch the football match. However, on both Saturdays a low and high peak at 17.00 and 18.00 were observed respectively. This time coincided with the end of the football match event as well as with the closure time of the shopping centre. Both the football stadium and shopping centre are located in the city centre. Hence, the post-match peaks on 28.04.18 were partly caused by the finish of football match and partly by the closure time of the shopping centre.

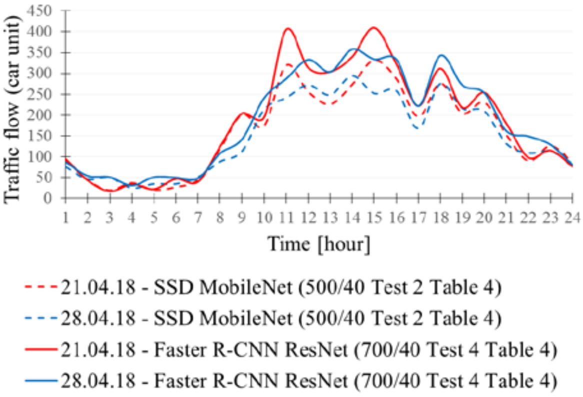


Figure 4. Car unit estimates on Saturday 28.04.18 date of a football match event and on Saturday 21.04.18 from a NECA CCTV location (NECA, 2018c).

Regarding the performance of two fine-tuned models, SSD MobileNet V1 COCO underestimated the car units with a maximum difference from Faster R-CNN ResNet 101 COCO of 82 units at 15.00 and 86 units at 11.00 on 21.04.18 and 28.04.18 respectively. Larger offsets between models were observed during day time when car estimates were higher than night time. This is because SSD MobileNet V1 failed to precisely distinguish each single car in images with a high car density within a close vicinity (i.e. when the cars are stand-still and roads are congested).

Traffic patterns before and during a snow event on 26-28th of February as well as during warm and dry weather conditions on 9-11th of July are depicted in Figure 5. The number of cars was signicifcantly reduced from Monday to Wednesday in February during rush hours, due to snowy conditions. It should be noted that the snow event commenced Monday night and continued the following days. Few observations in the early morning and late evening were unavailable, as the CCTV sensor was shut down. This is a common issue with CCTV technology during extreme weather conditions. Due to snow, a flat traffic pattern on Wednesday in February was observed, as opposed to the pattern on the same day in July with drier weather conditions. In comparison, high peaks in the early and late afternoons on 9 - 11th of July indicate the traffic pattern of standard working hours. Hence, the snow event clearly disrupted the typical daily traffic behaviour.

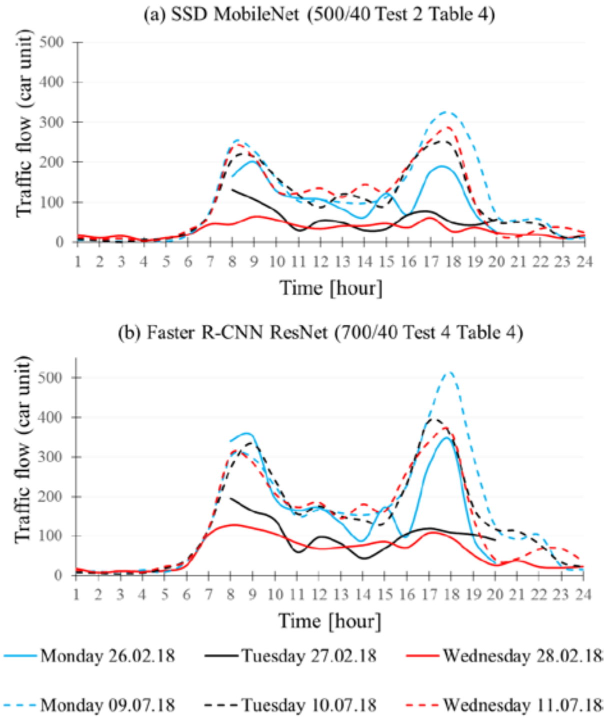
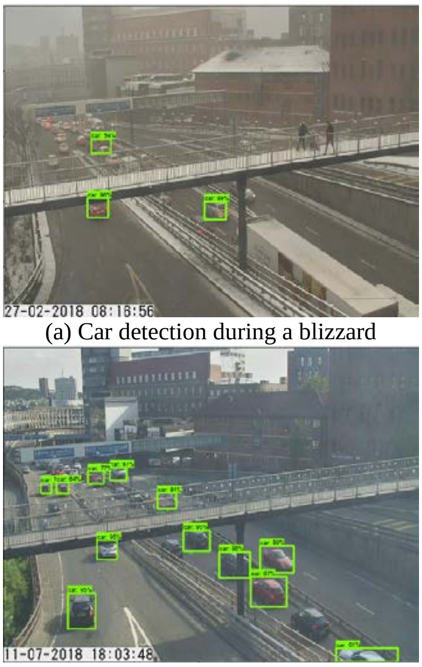


Figure 5. Car unit estimates on three consecutive days in February (a) before and during a snow event and in July (b) during dry warm weather from a NECA CCTV location (NECA, 2018a).

A considerably large difference in number of car unit estimates is observed during rush hours between results from SSD MobileNet and Faster R-CNN in Figure 5a and b respectively. Similar to previous traffic analysis shown in Figure 4, SSD MobileNet underestimated the number of cars. In particular, the snowy conditions adversely affected object detection performance, as small cars in the background of the image, seen in Figure 6a, could not be identified. It is also noticeable that lower than 100 cars were estimated on Wednesday 28.02.18 in Figure 5a. However, the perfomance of SSD MobileNet was improved on images captured on dry and sunny days, as evidenced in Figure 6b.

Nonetheless, both fine-tuned models provided similar flow trends with high peaks in the early morning and late afternoon, as also evidenced in Figure 7. These peaks were repetitive across all week days on 9-17th of July. Whereas a different traffic pattern was observed during the weekend on 14-15th of July. High peaks were apparent during midday but with approximately 200 fewer cars than those estimated on working days. Regarding the performance of the two models, fine-tuned Faster R-CNN outperformed in identifying more cars. An average difference of 38 car units across the nine-day monitoring period was calculated between SSD MobileNet and Faster R-CNN.



(a) Car detection during a blizzard

(b) Car detection during rush hour on a summer day

Figure 6. Car detection example with fine-tuned SSD MobileNet from a NECA CCTV location (NECA, 2018a).

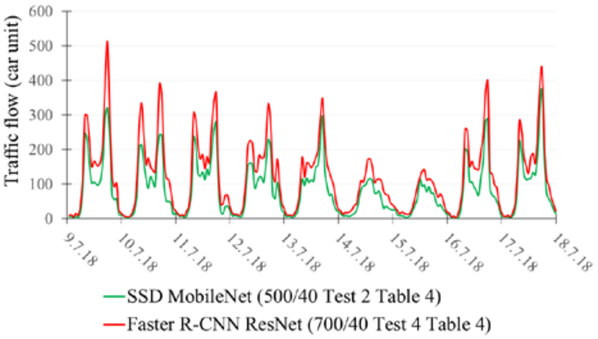


Figure 7. Car unit estimates on 9-17th July 2018.

4. DISCUSSION

Results from first experimental set have highlighted the significance to fine-tune pre-trained deep learning neural network systems with domain-specific imagery (i.e. NECA and UO CCTV images). Outcomes from the second experimental set have shown that fine-tuned Faster R-CNN performs better than SSD MobileNet model even after fine-tuning. This outcome complies with reported findings in a recent study Huang et al. (2017). Neural network systems performance can be improved after fine-tuning with larger image dataset and longer training.

In relation to speed, SSD MobileNet model works faster than Faster R-CNN. This also agrees with reports from previous studies (Liu et al., 2016; Gauen et al., 2017). However, SSD MobileNet provides limited performance in detecting small cars and cars on noisy images. This is possible due to SSD's receptive field being smaller than R-CNN's. It should be noted that at all experiments configuration settings for training were kept unchanged and retrieved from Rathod et al. (2018). Future tests with varying parameterisation (e.g. image input size) can potentially improve SSD's performance.

Examples of car unit estimates have shown the variations of traffic patterns on hourly and daily-basis. This analysis can ultimately support the understanding of traffic disruption and congestion before, during and after an extreme weather condition event or a social event. It can be suggested that the fine-tuned MobileNet model with 98.2% precision, 58.5% recall and 73.4% harmonic mean could potentially be used for a real time traffic analysis with big data. This is because it provides a trade-off between fast and precise performance (Model A Test 2 in Table 4). Whereas, Faster R-CNN (Model C Test 4 in Table 4), with a better harmonic mean and recall than MobileNet, of 80.4% and 68.8% respectively, could be used for traffic analysis applications that demand higher accuracy than speed (e.g. urban planning and management of road traffic infrastructure).

To better interpret traffic flow behaviour from CCTV datasets with deep learning technology, additional parameters should be taken into consideration, such as a) environmental conditions as they can introduce noise on imagery, and b) the performance of each neural network system in terms of precision and speed even after fine-tuning. Similar traffic analysis is scheduled from CCTV sensors of multiple key locations that link the city centre with its outskirts. Additional tests will estimate the direction of moving cars in relation to the city centre. This could further support decision making for stakeholders in case of evacuation after prediction of an extreme weather event.

5. CONCLUSIONS

Investigations in the presented research have demonstrated the capabilities of state-of-the-art supervised deep learning algorithms in car detection from CCTV big data for urban traffic monitoring with low quality datasets. Tests were performed with imagery captured on days during varying environmental conditions and seasons of the year 2018, from two CCTV locations close to Newcastle upon Tyne city centre. Fine-tuning, with domain specific imagery, of available to-the-public pre-trained neural network systems, such as SSD MobileNet V1 COCO and Faster R-CNN ResNet 101 COCO, optimised traffic flow estimation during disruptive events. Traffic flow analysis is fundamental for urban planning and management of road traffic infrastructure.

ACKNOWLEDGMENTS

This research was funded by the Natural Environment Research Council (NERC) Flood-PREPARED award (NE/P017134/1; https://www.ncl.ac.uk/engineering/research/civil/water/projects/flood-prepared.html). The North East Combined Authority (NECA) Travel and Transport Data are used and published in this paper with the permission of the Open Government License 3.0, Newcastle City Council, 2014, Newcastle upon Tyne, UK (https://www.netraveldata.co.uk/?page_id=13).

REFERENCES

Buch, N., Velastin, S. A. and Orwell, J. 2011. A review of computer vision techniques for the analysis of urban traffic. *IEEE Transactions on Intelligent Transportation Systems*, 12(3)**,** pp. 920-939, doi:10.1109/TITS.2011.2119372.

Gauen, K., Dailey, R., Laiman, J., Zi, Y., Asokan, N., Lu, Y. H., Thiruvathukal, G. K., Shyu, M. L. and Chen, S. C. 2017. Comparison of visual datasets for machine learning. Proceedings - 2017 IEEE International Conference on Information Reuse and Integration, IRI 2017, 2017. pp. 346-355.

Geiger, A., Lenz, P. and Urtasun, R. 2012. Are we ready for autonomous driving? The KITTI vision benchmark suite (http://www.cvlibs.net/datasets/kitti/). 2012 IEEE Conference on Computer Vision and Pattern Recognition, 16-21 June 2012 2012. pp. 3354-3361.

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. 2017. *MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications*, Computer Research Repository (CoRR), Cornell University (https://arxiv.org/abs/1704.04861).

Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S. and Murphy, K. 2017. Speed/accuracy trade-offs for modern convolutional object detectors. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017. pp. 3296-3305.

James, P. M., Dawson, R. J., Harris, N. and Joncyzk, J. 2014. Urban Observatory Weather and Climate data. Newcastle University, doi:10.17634/154300-20.

Krizhevsky, A., Sutskever, I. and Hinton, G. E. 2012. ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 2012. pp. 1097-1105.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W. and Jackel, L. D. 1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4)**,** pp. 541-551, doi:10.1162/neco.1989.1.4.541.

Li, Y. 2008. Short-term prediction of motorway travel time using ANPR and loop data. *Journal of Forecasting*, 27(6)**,** pp. 507-517, doi:10.1002/for.1070.

Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C. L. 2014. Microsoft COCO: Common objects in context. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* doi:10.1007/978-3-319-10602-1_48.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y. and Berg, A. C. 2016. SSD: Single shot multibox detector. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* doi:10.1007/978-3-319-46448-0_2.

Lv, Y., Duan, Y., Kang, W., Li, Z. and Wang, F. Y. 2015. Traffic Flow Prediction with Big Data: A Deep Learning Approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2)**,** pp. 865-873, doi:10.1109/TITS.2014.2345663.

NECA, 2018a. *CCTV sensor overlooking the A167 Central Motorway from the A193 Newbridge Street Roundabout* [Online]. Available: https://netrafficcams.co.uk/newcastle-a167-central-motorway-a193-newbridge-street-roundabout-0 [Accessed: 24 May 2018].

NECA, 2018b. *NECA CCTV database* [Online]. Available: https://netrafficcams.co.uk/ [Accessed: 2 February 2018].

NECA, 2018c. *NECA CCTV sensor located at A167 High Street, Tyne Bridge, Gateshead* [Online]. Available: https://www.netrafficcams.co.uk/gateshead-a167-high-street-tyne-bridge-0 [Accessed: 15 May 2018].

Rathod, V., Pkulzc and Wu, N., 2018. *Raw pre-trained TensorFlow models for various neural network systems* [Online]. Available: https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/detection_model_zoo.md [Accessed: 15 June 2018].

Ren, S., He, K., Girshick, R. and Sun, J. 2017. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 39(6)**,** pp. 1137-1149, doi:10.1109/TPAMI.2016.2577031.

Shi, S., Wang, Q., Xu, P. and Chu, X. 2017. Benchmarking state-of-the-art deep learning software tools. Proceedings - 2016 7th International Conference on Cloud Computing and Big Data, CCBD 2016, 2017. pp. 99-104.

Tang, T., Zhou, S., Deng, Z., Lei, L. and Zou, H. 2017. Arbitrary-oriented vehicle detection in aerial imagery with single convolutional neural networks. *Remote Sensing*, 9(11), doi:10.3390/rs9111170.

TensorBoard, 2018. *TensorBoard: Visualising Deep Learning* [Online]. Available: https://www.tensorflow.org/guide/summaries_and_tensorboard [Accessed: 15 February 2018].

UO, 2018. *Urban Observatory, weather, climate and traffic data, generated by Newcastle University* [Online]. Available: http://uoweb1.ncl.ac.uk/ [Accessed: 30 June 2018].

Wang, X., Zhang, W., Wu, X., Xiao, L., Qian, Y. and Fang, Z. 2017. Real-time vehicle type classification with deep convolutional neural networks. *Journal of Real-Time Image Processing*, doi:10.1007/s11554-017-0712-5.