Recognition of Traffic Congestion State using a Deep Convolutional Network

C. J. Willis
BAE SYSTEMS AI Labs
Great Baddow CM2 8HN, UK
chris.willis@baesystems.com

D. Harborne
University of Cardiff
Cardiff CF10 3AE, UK
HarborneD@cardiff.ac.uk

R. Tomsett
IBM UK Ltd.
Hursley SO21 2JN, UK
RTomsett@uk.ibm.com

M. Alzantot
CS Dept., UCLA
Los Angeles, CA 90095, USA
malzantot@ucla.edu

***Abstract*—This paper describes a two-phase deep learned neural network for the recognition of traffic congestion in surveillance camera imagery. The network is transfer-learned using GoogLeNet for image processing and a bespoke subnet for congestion recognition.**

I. Introduction

Monitoring traffic congestion within a region provides an interesting and challenging problem in multi-modal, temporal situational awareness. The approach examined in this paper addresses the problem of assessing the state or level of congestion at road junctions using only imagery collected at that location. This assessment will form one input to a higher-level situational awareness capability that takes account of multiple data sources.

II. Image Data

Transport for London (TfL) manages a network of surveillance cameras which are used to monitor road junctions in the city [1]. Images and video sequences (of a few seconds duration) from over 1,200 cameras are collected and made available for third party use in the development of new applications. The images and video sequences are released with a five-minute refresh rate. This paper focuses on processing the still images; these are 352×288 pixels.

A data collection campaign was organized which collected both images and video sequences from 691 cameras over a period of 23 days. This set includes imagery at all times of the day and night, in different location types and in the range of traffic and environmental conditions prevailing at collection time. Typical examples of images collected and used in this paper are shown in figure 2. These images illustrate just some of the variation in lighting, traffic and road configuration which is present in the dataset. The archive will ultimately be made available for researchers working with surveillance imagery to exploit.

III. Labelling

A subset of the imagery covering five locations and 24 hours was selected for ground truth labelling. In order to carry out this step an internet accessible image annotation tool [2] has been developed which allows the labelling of each image as one of: uncongested, congested, unknown or broken. The former pair are for use on uncontentious imagery from the cameras. The unknown type tends to be selected for borderline cases in which it is difficult to say whether traffic is congested or not. The final 'broken' type is recommended for imagery from cameras deployed on activities other than traffic monitoring where a test-card message is typically displayed, or for imagery in which the camera is not directed at road junctions. The annotation tools is made available to researchers working on the task for crowd-sourced truthing activities.

At the time at which the data for this paper was downloaded a total of 4,117 images had been labelled by at least one annotator. The broken images were rejected, leaving a total of 3,967 usable labelled examples. Taking the most prevalent labelling for each image revealed that 57% were uncongested and 43% congested. In order to take account of the non-binary nature of the phenomena of interest, and differing opinions expressed when labelling the imagery, soft class memberships are derived for use in training and validation. These take into account the labels supporting a class, against the class and for the 'unknown' labels:

$$p_i = \frac{v_i + u + 1}{\sum_{j=1}^{N}(v_j + u + 1)}.$$

Here the v_i are the number of annotations assigned to class i and u are the number of annotations for the unknown class. This type of soft class membership avoids some issues caused by the hard classification of borderline cases adversely biasing results. For classifier development the images were randomly partitioned into a training set of 80% of images and 10% each for of the validation and test sets.

IV. Classifier Development

The training, validation and test datasets themselves were generated by collecting a number of 200×200 pixel sub-images from each starting image. Firstly the central sub-image is collected, then the eight overlapping images formed by offsetting the selection region by 25 pixels to the left, right, up and down. This is repeated for left-right flipped variants so that an augmented dataset of over 57,000 training and 7,000 validation and test images respectively. The sub-images derived largely avoid the TfL-furnished annotation which lies towards the edges of the originals.

The congestion classifier uses a pair of deep neural networks. The first uses the GoogLeNet [3] Inception network,

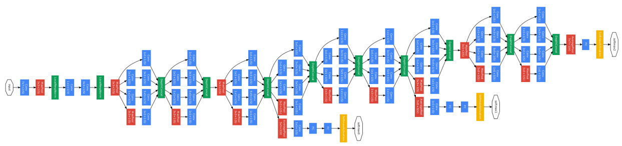

Fig. 1. GoogLeNet Architecture

pre-trained on ImageNet data, for the image processing feature extraction step (figure 1). A feature vector for transfer learning is tapped-off before the network's fully connected layer, between the rightmost average pooling layer (red) and the fully connected layer (blue). The feature vector is passed into a five-layer fully-connected network to transform the image features into a classification assessment. This network uses hyperbolic-tangent and softmax activation functions and is trained using the Adam optimizer until the cross-entropy on the validation set stops improving. Both networks are implemented in TensorFlow [4].

V. Results

A selection of classified images is shown in figure 2. The figure shows an image classified as uncongested at the top, a borderline case in the middle and an image classified a congested at the bottom. Performance on the test set as a whole indicates that the proportion of images classified according to most prevalent label is over 95%, and examination of the misclassified images revealed that most were on the classification borderline.

VI. Conclusion

A deep-learned classifier for street-camera imagery has been developed for the recognition of congestion at road junctions. The classifier is trained on hand-labelled open-source imagery of London's road junctions and returns a high-performance given the subjective nature of the problem. It has proved robust to different road configurations, times-of-day, lighting, atmospheric conditions and other distractors.

The classifier is designed to be one of several inputs to a higher-level capability which combines data from different sources and modalities for the extraction of situational awareness. Future work will further our exploration of the coalition situational understanding problem space. The most immediate step for this will be to add interpretability to our classification model. This will help tackle the issue of trust across the coalition when sharing intelligence and can help share justifications for decisions without having to release restricted input data.

Acknowledgment

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence under Agreement Number W911NF-16-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defence or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation hereon.

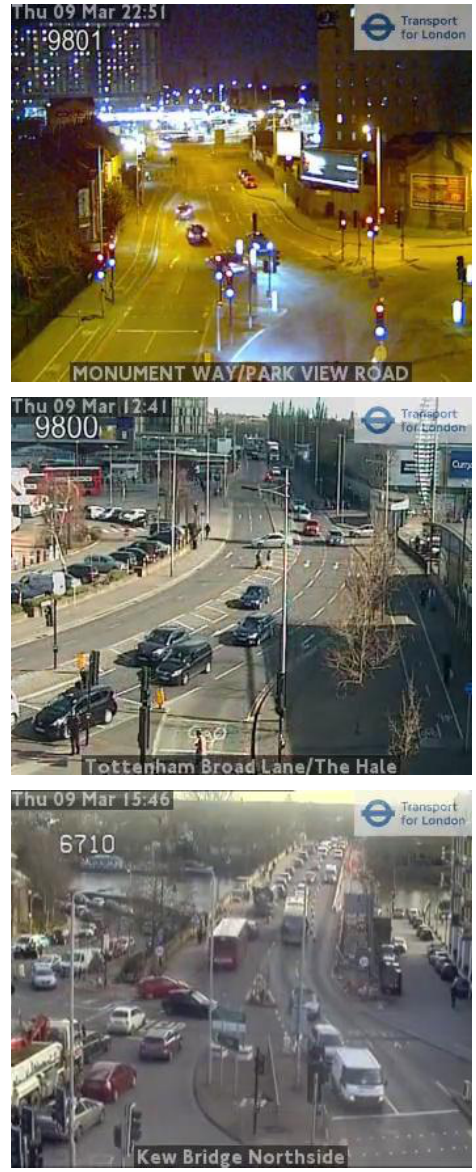

Fig. 2. Classified Images: uncongested (top), borderline (middle) and congested (bottom)

References

[1] Transport for London, London Jam Cams, http://www.trafficdelays.co.uk/london-traffic-cameras/, last visited June 2017.

[2] Tomsett, R., Congestion Image Annotation Tool, https://image-annotation.eu-gb.mybluemix.net/, last visited June 2017.

[3] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., *Going Deeper with Convolutions*, Computer Vision and Pattern Recognition (CVPR), http://arxiv.org/abs/1409.4842, 2015.

[4] Abadi. M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., & 34 others, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*, http://tensorflow.org/, 2015.