Distributed Opportunistic Sensing and Fusion for Traffic Congestion Detection

Alistair Nottle*, Daniel Harborne†, Dave Braines§†, Moustafa Alzantot‡, Santiago Quintana-Amate*
Richard Tomsett§, Lance Kaplan¶, Mani B. Srivastava‡, Supriyo Chakraborty‖, Alun Preece†
*Airbus Central R&T, †Crime and Security Research Institute, Cardiff University, ‡UCLA,
§IBM UK, ¶US Army Research Laboratory, ‖IBM US

***Abstract*—Our particular research in the Distributed Analytics and Information Science International Technology Alliance (DAIS ITA) is focused on "Anticipatory Situational Understanding for Coalitions".**

This paper takes the concrete example of detecting and predicting traffic congestion in the UK road transport network from existing generic sensing sources, such as real-time CCTV imagery and video, which are publicly available for this purpose. This scenario has been chosen carefully as we believe that in a typical city, all data relevant to transport network congestion information is not generally available from a single unified source, and that different organizations in the city (e.g. the weather office, the police force, the general public, etc.) have their own different sensors which can provide information potentially relevant to the traffic congestion problem. In this paper we are looking at the problem of (a) identifying congestion using cameras that, for example, the police department may have access to, and (b) fusing that with other data from other agencies in order to (c) augment any base data provided by the official transportation department feeds. By taking this coalition approach this requires using standard cameras to do different supplementary tasks like car counting, and in this paper we examine how well those tasks can be done with RNN/CNN, and other distributed machine learning processes.

In this paper we provide details of an initial four-layer architecture and potential tooling to enable rapid formation of human/machine hybrid teams in this setting, with a focus on opportunistic and distributed processing of the data at the edge of the network. In future work we plan to integrate additional data-sources to further augment the core imagery data.

I. Introduction

Congestion is an indicator that a transport network is either routinely not fully meeting the needs of the users (i.e. there is not enough capacity to meet regular demand), or it is being disrupted by an extraordinary occurrence such as an accident or major public event. Transport network users and those responsible for maintaining the network desire increasingly fine-grained real-time visibility of the evolving network based on actual usage, and the potential for harvesting this data, in an ambient manner, from distributed devices in addition to bespoke ancillary systems.

Multiple data sources can play a role in detecting congestion, spanning multiple modalities, such as traffic cameras, public transport data, weather *etc*. However, the sources may not all be under a single authority or organisation, requiring establishment of a 'coalition' of related sources and information processing to address the problem of detecting congestion. We term this aspect of the coalition context as a 'coalition of convenient sources'. By accounting for such a 'coalition of sources' in a situational awareness solution there is much opportunity to enhance the situational picture and go beyond perception into comprehension and projection. For example, if one camera indicates that a particular street corner is congested, why is it congested? Will the congestion spread? How can a coalition take prescriptive action to reduce the congestion or to ensure that the congestion does not spread?

Developing situational understanding entails efficient usage of both human and machine agents working together as a hybrid team as seamlessly as possible, within a coalition setting. To support the situational understanding requirement we are specifically investigating machine learning and deep learning techniques for the machine agents with the human agents acting as consumers of the information arising, as well as being able to modify the behaviour of the system based on the contribution of contextually relevant human knowledge. The ability for the human users to interact with the machine agents is a key focus of our work, both in terms of the ability to configure and direct the machine processes, as well as the ability for the machine agents to explain themselves and their processing to humans to aid interpretation.

This paper first provides an overview of the current state-of-the art for detection and measurement of traffic congestion. To advance our 'coalition of convenient sources' perspective, a multimodal dataset has been gathered using publicly accessible temporally and spatially coordinated data feeds which are distributed both geographically and organisationally, i.e. the data sources converge on a common geographic region but could plausibly be provided by different coalition partners. Some techniques for the analysis of this data are proposed and compared, specifically in terms of detecting or inferring traffic congestion in real-time. A high-level architecture is proposed for how these distributed sensors, as well as underlying uncertainty, could be fused to provide a holistic map of congestion across the sensed network and potentially beyond. This component-based architecture accounts for current capabilities but also enables new relevant capabilities to be easily inserted as they become available in the future.

II. Summary of Prior Art

Addressing the problem of detecting traffic congestion is not a new challenge. With the recent explosion in sensor availability and a drive towards 'smarter cities' both this body of knowledge and the associated data sources will increase [1]. An obvious starting point are the metrics used to determine congestion from simple determinations such as stream speed [2] to more complex characterisations such as a Traffic Congestability Value (TCV) [3], as well as how congestion propagates across the network [4] and designing networks to be congestion resistant.

There are some investigations into multiple modalities for congestion sensing from more indirect measures such as weather [5] and social media [6], [7], as well as more direct observations of the traffic network, such as the use of computer vision for analysing images for signs of congestion [8], [9].

Congestion determinations can also be part of a feedback loop into smart or autonomous control systems [8], [10], [11], thereby enabling the action taken to affect the sensed stimulus.

Our research in this space is focused specifically on opportunistic, distributed sensing, in particular alongside the use of *machine learning / deep learning* tools, in scenarios where common computational modes or tooling may not be shared among all partners. Of clear relevance to this more distributed and opportunistic approach are some interesting studies using more traditional toolsets on existing map service providers to provide wider indications of congestion [12], [13], as well as other facets such as SVM [14], HMM Models [15], multi-agent based models [11], long short-term memory [13], and backpropagation (BP) Neural Networks [16].

Some have investigated the use of connected vehicles as a distributed sensing and computation platform. Local networks of connected vehicles can make use of vehicle ad-hoc networks (VANETs) to share a local picture of congestion in a distributed fashion without relying on underlying or existing infrastructure [17], as well as cooperating to minimise excessive use of constrained resources [18].

III. Description of Datasets

To support the development and evaluation of the proposed techniques a multi-modal dataset has been gathered bringing together temporally and geographically aligned data from public datasources. London (UK) was chosen as there were a number of appropriate data feeds which were accessible to researchers, mainly via the Transport for London (TfL) API [19] which provides access to traffic cameras (still and video), live bus updates, air quality, cycle data, and more. The CCTV image and video data is collected every 5 minutes with a 10 second video clip and a single still image being available for each of the approximately 1000 CCTV cameras with a publicly available live feed. Each day of collection sees around 33GB of video data, 5GB of still images, and 750MB of other sensor data captured, totalling around 40GB. This has been gathered continuously since February 2017 in order to create a consistent dataset for machine learning and pattern-of-life model building.

Whilst much of the data comes from an authoritative source (TfL), there are still gaps and inconsistencies which arise sometimes owing to technical faults (e.g. camera or camera feed issues), and at other times data is intentionally withheld for security purposes. For instance, in the TfL CCTV data, camera imagery may be replaced with a static holding image stating the camera is being used to 'keep London moving'. In a coalition environment, similar occurrences may arise where one partner may want to intentionally withhold data due to sensitive operations in that area or other security or privacy related reasons.

A. Labelling of image data

An in-house web service has been developed to enable rapid subjective labelling of the image data with a 'ground truth' estimation of the levels of congestion. This is achieved through crowd-sourcing opinions on whether images from traffic cameras show congestion. A day's worth of images from a subset of cameras, selected by highest availability, were uploaded to the tool, and then randomly presented to participants. Through simple assigned keystrokes, participants marked the image as either 'No Congestion', 'Congestion', 'Don't Know', or 'Broken Image'.

To enable the human users crowd-sourcing the classifications a working definition of congestion was formulated as '*a line of traffic that would cause you to slow or stop*'. This phrasing was chosen deliberately to convey exactly what the human users should treat as congestion regardless of their individual background or experiences. For example, people from large cities would likely have a different subjective definition of congestion to those living in a more rural setting. The participants were encouraged to make a subjective assessment of the traffic levels they see directly in the image rather than to try to 'second guess' any possible machine learning system. For example, the human users were encouraged to take into account subtle visual clues such as the gaps between vehicles (to suggest high speed or low speed movement), or the presence of traffic lights. The human users were discouraged from trying to guess how the machine agent might classify an image but instead focus purely on their human reaction.

IV. Proposed Analysis Tools

Not all the multi-modal inputs can be readily fused to determine congestion. A lot of value resides in the imagery data in particular, and work has been done investigating how existing deep learning techniques can be applied to enhance other image processing techniques to understand the congestion indicated directly in the static images and video.

A. Image Segmentation to identify occupancy

Contrasting approaches of a VGG-16 model trained for single object classification purposes against a Fully Convolutional Network [20] were investigated to see their suitability for classifying congestion from CCTV images. Initially an image classification task was performed to see if using a pre-trained convolutional neural network (obtained using 'ImageNet' dataset) it is possible to identify all the objects present in an image gathered from a CCTV camera. The pre-trained model was slightly modified in order to be applied in a fully convolutional procedure as described in [20].

The preliminary results obtained were not very promising, mainly since the original model was created to carry out classification of a single object. Images within the dataset often contained multiple vehicles with occlusions and overlaps. To overcome this issue a posterior experiment was performed where the fully convolutional VGG-16 model [21] was trained on a segmentation dataset ('PASCAL VOC'). However, no improvements to prediction accuracy were obtained. Lack of clear measures of the number of objects in the image make it difficult to use occupancy as a measure of congestion. The low accuracy results are likely to be caused by the use of a completely different dataset for the training phase. Other circumstances may also have affected the accuracy of the predictions such as the relatively low-quality images from the CCTV cameras, the small size of some of the vehicles in the image, the density of vehicles in the congested images, and frequent occlusions to the scenes.

B. Car Detection / Counting using R-CNN

As discussed in the previous section, the classification approach is insufficient for detecting congestion from the provided images. To address this problem, we use regional-convolutional neural network (R-CNN). R-CNN can be used to localize and detect *multiple* objects within the input image. While the same thing can, in theory, be achieved by using CNN to classify different sliding windows inside the given image, this approach is too computationally expensive. R-CNN provides a computationally efficient solution for localizing and detecting objects in the image. Fine-tuning of the pre-trained model using images sampled from our dataset is planned as well as modification of the model architecture to keep only output classes that are valid for the congestion detection problem.

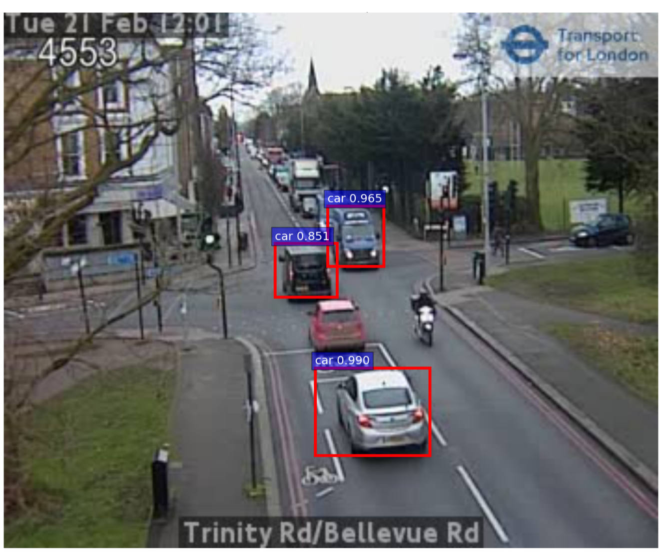

Fig. 1. Example of car detection using R-CNN

1) Results: We use TensorFlow [22] to evaluate the Faster R-CNN algorithm [23] on example images drawn from our dataset (see figure 1). We used a model pre-trained on the Pascal VOC2007 [24] dataset. We ran our experiments using a desktop machine equipped with NVIDIA TitanX GPU to accelerate the model running time.

2) Discussion: Although our results show that R-CNN can detect and localize cars in the images with high precision (*low false positives rate*), we find that the model suffers from low recall (*high false negatives rate*). In order to improve our model, we propose the following refinements:

- Fine-tune the model using a training subset from our dataset. Although using pre-trained model weights trained on large dataset such as ImageNet/PASCAL VOC is useful when having another relatively small labelled datasets like ours. The discrepancies between the two datasets would require that we fine-tune the model using data sampled from our dataset.
- Replace the final fully connected layer in the fine-tuned model with another fully connected layer with a smaller number of output classes to remove classes that are not applicable in our scenario (e.g. cat, TV screen, etc.). This will result in both higher accuracy and faster running time of our model.

V. High Level Architecture

In order to manage and fuse these sensor feeds, a high level architecture is proposed, consisting of four hierarchical layers. Each of these layers is virtual and can span multiple agencies within a coalition.

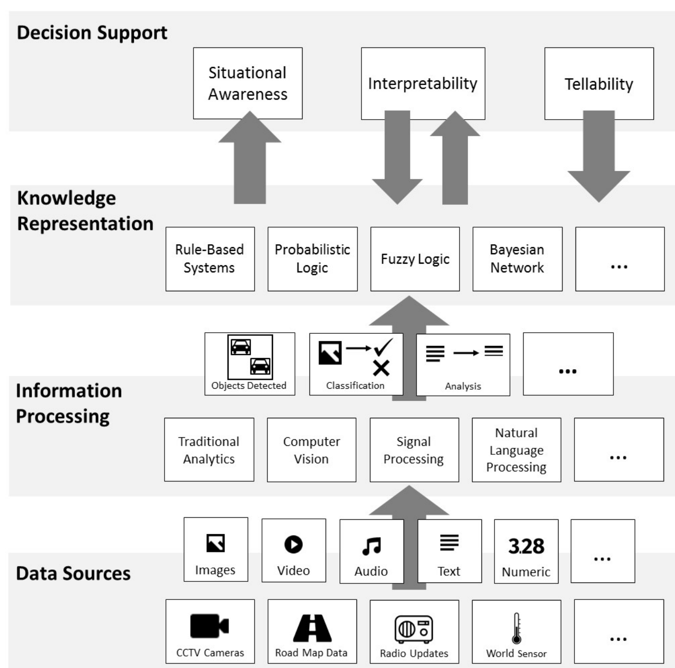

Fig. 2. Proposed High Level Architecture

As outlined, we are looking at the traffic congestion problem as a setting to explore the coalition situational understanding problem space. To do this, we have developed a conceptual system architecture that defines and explores the relationships between data sources and services owned by various partners of a coalition. Ultimately, the architecture outlines the ways in which a coalitions resources can be used to provide actionable intelligence and decision making assistance.

In the following sections, we will begin by providing a summary of the architecture and then present how its use can deal with some of the challenges of the coalition environment.

A. Data Sources Layer

The *data sources* layer is made up of heterogeneous data sources provided by a range of sensors, modalities, and collection platforms. Ownership for each source may lie with different members of the coalition who grant access as required, or any source could be publicly available and therefore open source in nature. Many additional capabilities and complexities can be inserted into this layer of the architecture, for example through using fine-grained policy based access control to allow each member of the coalition to share their sources explicitly with certain other partners only.

B. Information Processing Layer

At the *information processing* layer, processing services are maintained and shared by the partners of the coalition. These services are responsible for producing initial conclusions from the input data provided by the coalition data sources. For example, identifying objects within images. This initial information may come with a level of uncertainty which will be recorded as metadata and can then be passed through the system, enabling it to be available to higher level services and users to be taken into consideration when creating combined data products from these initial results. This enables the inherent uncertainty arising from this basic processing to be explicitly captured and bound to the resulting information to provide additional accountability and transparency in any subsequent usage, and to better enable the interpretability of results later in the pipeline.

C. Knowledge Representation Layer

The *knowledge representation* layer contains all processes relating to the semantics or meaning of the data. Meaning is almost always contextual, and the domain models which can help define this processing context can be defined in this layer. There are many different types of processing that can be carried out in the knowledge representation layer and it is here that disjoint information coming out of the lower information processing layer can be fused or related. This is also the layer where the human users have the most opportunity to inject their human knowledge into the system, enabled via the tellability function (outlined later) in the uppermost layer. Different models and techniques can be used in the knowledge representation layer to fulfil different kinds of processing and many of these can be integrated to enable high-value outcomes. For example, current work into the integration of Subjective Logic and Bayesian Networks is showing that a usable Bayesian Network can be constructed with far less training data if subjective human opinions can be encapsulated in the links [25].

D. Decision Support Layer

The final layer, *decision support*, allows the user agent to utilise the resources in the layers below. Unlike the lower layers, which consists of services distributed across the coalition, each agent would have their own instance of a user interface which would enable querying of the available services to assist in the tasks of the agent.

We have outlined three major functions of the decision support layer. The first, *situational awareness*, is the general ability of the system to assist the agent in making decisions through their awareness of the environment.

The second, '*interpretability*', outlines the ways in which the user can query the system for reasons and explanations that back the intelligence and recommendations being given by the system. The interpretability response from the system may vary in strength and modality dependant on the services below that have been involved in the reasoning process, and may also be mediated by the uncertainty encapsulated within the metadata.

Finally, the decision support layer offers '*tellability*' - the ability for a user agent to inject knowledge in to the services below in order to correct or improve the reasoning generated by the system. This is a key differentiator in this architecture which enables the human users to directly affect the behaviour of the system through the addition of human knowledge. The simplest case is knowledge which can be directly used by processes within the knowledge representation layer, for example, to perform improved reasoning or inference, or to modify the weights in a Bayesian Network. More complex system changes can be affected in the lower layers by the additional human knowledge causing different configuration parameters to be used or entirely different information processing components to be used. Finally, the human user may tell the system about new data sources, thereby increasing the available pool of data sources at the lowest level of the architecture.

VI. Benefits to a Coalition Environment

A. Computational Resource Efficiency

By having layers of services as outlined, each layer can optimise the number of calls made to services below in order to increase efficiency or to adhere to any usage limits imposed by the service's owning partner. As a concrete example, we have used a combination of an optical flow blob detection service combined with the R-CNN car detector mentioned previously to identify moving cars in TfL camera video.

The R-CNN is more computationally expensive than the optical flow algorithm and thus, to optimise the efficiency of reasoning with the two components, calls are made every frame to the optical flow algorithm but only made to the R-CNN after an interval of frames.

Without the calls to the R-CNN the information generated is reduced only to 'blobs in motion' and thus can't be used to reason about the traffic conditions since the blobs are not identified as cars (they could be pedestrians, birds, litter, *etc*). With the periodic calls to the R-CNN the system gains the classification of certain features as being detected cars for a given frame with no concept of motion. By taking the 'blobs in motion' and the fact that in one frame of the video a given blob was identified as a car we can therefore infer 'cars in motion' but only when fusing the information from these two services together can the system perform the required reasoning in the context of a traffic congestion scenario.

B. Load Sharing & Redundancy Through Failure and Policy Change

Given the nature of the proposed architecture, the decision support layer can benefit from multiple services performing similar tasks creating redundancy and the ability to optimise the load put on each of the redundant services.

In our work, we are exploring using both an LSTM ('Long Short-Term Memory') based service and HTM ('Hierarchical Temporal Memory') based service to anticipate future traffic levels. This has allowed us to explore and demonstrate that even given coalition constraints, such as infrastructure limitations or modification to information sharing policy from a coalition partner, the system is still able to provide assistance to user agents as long as one service is still available. In the case where both services are available the system can balance the load between them if only one service is required rather than using both for cross validation (although that use-case is equally valid, albeit not from the perspective of reducing system load).

C. Diversification of Sources and Strength of Reasoning

An alternative application of multiple services with similar input/output signatures is to use both for reasoning in order to provide cross validation and reduce uncertainty.

As highlighted earlier, the information produced by the information processing layer can come with a level of uncertainty (for example the confidence a model has in its classification). By having services in the knowledge representation layer that make use of multiple sources of information confidence levels can be balanced to strengthen the intelligence being provided to the end user agent. In addition, the use of a diverse information service offering allows the knowledge representation services to 'learn' the reliability of the information providers which can create an inherent rating of trust for each service in the layer below.

We are currently using our developments of subjective Bayesian networks to explore this trust aspect of the framework [25].

D. Knowledge representation models and techniques

The exact details of the knowledge representation models and techniques are still under active definition within our architecture (in terms of specific languages, logical expressivity levels, candidate components *etc*), however the functionality to be provided by this layer is clear. Figure 3 shows the operational context of a simplified sub-section of our overall approach (described previously) and shows the meta information sitting as a key resource between the human users and the various machine components within the system. This meta-information encapsulates the knowledge representation models and is used to drive the behaviour of the machine components, the selection of data sources, the higher level reasoning (if needed) and as a source of more information provided by the human users as described previously. The role of the DL networks, the Bayesian network and the human user along with the various feedback loops is described in more detail in [26].

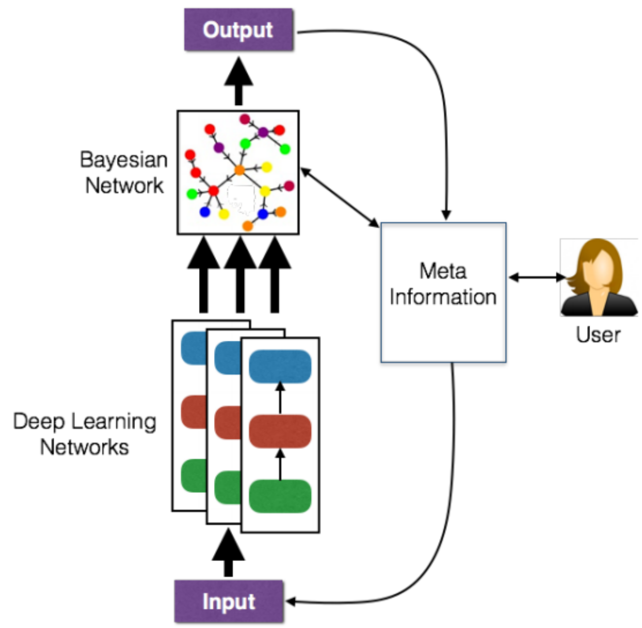

Fig. 3. Role of Knowledge Representation

Our aim in this part of the architecture is to provide a single human-friendly representation language that is easily configured and extended and is easy to parse by machine agents. This will be used to drive each of the individual reasoning components with translation to other representation formats as required. Candidates for this meta information representation language include: The Web Ontology Language (OWL) or Resource Description Framework (RDF), JSON or JSON-LD, a form of Controlled Natural Language (e.g. ITA Controlled English [27]). The key requirement is full flexibility and extensibility with a strong desire for ease of human use and efficiency of machine processing.

VII. Conclusion

Overall, our proposed architecture shows promise for this use case as well as having the potential to expand to other broader situational understanding scenarios. At this initial stage in our research we are exploring the overall architecture of a possible system like this to support rapidly formed coalitions of human-machine hybrid teams and have taken an initial deep dive into some of the machine learning and deep learning components with this traffic congestion scenario. In future work we plan to more deeply integrate the capabilities of the knowledge representation layer, enabling information about the data sources, the processing and relevant local knowledge from human users to be stored and acted upon. We believe that the integration of these components plus human input in real-time could provide the basis for a powerful and flexible system to support such coalition teams in the near future. We also anticipate that our component based approach gives a fair degree of future-proofing as new techniques and advances are made in this rapidly evolving field.

In order to prove the potential benefits of our approach we plan to run a series of experiments with each of the components within the architecture in order to determine their individual performance characteristics. This will provide a strong baseline for further potential system-level experiments to measure the capabilities of the entire architecture, especially in terms of rapid formation and re-configuration which would be a key benefit arising from this hybrid approach.

Acknowledgement

This research was sponsored by the U.S. Army Research Laboratory and the UK 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 UK Ministry of Defence or the UK Government. The U.S. and UK Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copy-right notation hereon.

References

[1] M. Naphade, G. Banavar, C. Harrison, J. Paraszczak, and R. Morris, "Smarter cities and their innovation challenges," *Computer*, vol. 44, no. 6, pp. 32–39, June 2011.

[2] A. M. RAO and K. R. RAO, "Identification of traffic congestion on urban arterials for heterogeneous traffic." *Identificacion de la congestion del trafico en las zonas urbanas para el trafico arterials heterogeous.*, vol. 11, no. 3, pp. 131–142, Jul. 2016.

[3] Patel Nilanchal and Mukherjee Alok Bhushan, "Assessment of network traffic congestion through Traffic Congestability Value (TCV): a new index," *Bulletin of Geography: Socio-Economic Series, Vol 30, Iss 30, Pp 123-134 (2015)*, no. 30, p. 123, 2015.

[4] H. Nguyen, W. Liu, and F. Chen, "Discovering Congestion Propagation Patterns in Spatio-Temporal Traffic Data," *IEEE Transactions on Big Data*, vol. PP, no. 99, pp. 1–1, 2016.

[5] J. Lee, B. Hong, K. Lee, and Y. J. Jang, "A Prediction Model of Traffic Congestion Using Weather Data," in *2015 IEEE International Conference on Data Science and Data Intensive Systems*, Dec. 2015, pp. 81–88.

[6] S. Wang, L. He, L. Stenneth, P. S. Yu, Z. Li, and Z. Huang, "Estimating Urban Traffic Congestions with Multi-sourced Data," in *2016 17th IEEE International Conference on Mobile Data Management (MDM)*, vol. 1, Jun. 2016, pp. 82–91.

[7] S. Wongcharoen and T. Senivongse, "Twitter analysis of road traffic congestion severity estimation," in *2016 13th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, Jul. 2016, pp. 1–6.

[8] M. R. Islam, N. I. Shahid, D. T. u. Karim, A. A. Mamun, and M. K. Rhaman, "An efficient algorithm for detecting traffic congestion and a framework for smart traffic control system," in *2016 18th International Conference on Advanced Communication Technology (ICACT)*, Jan. 2016, pp. 1–1.

[9] I. Ali, A. Malik, W. Ahmed, and S. A. Khan, "Real-Time Vehicle Recognition and Improved Traffic Congestion Resolution," in *2015 13th International Conference on Frontiers of Information Technology (FIT)*, Dec. 2015, pp. 228–233.

[10] S. Surya and N. Rakesh, "Flow based traffic congestion prediction and intelligent signalling using Markov decision process," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, vol. 3, Aug. 2016, pp. 1–6.

[11] A. Lakhouili, E. H. Essoufi, and H. Medromi, "Multiagent based model for urban traffic congestion measuring," in *2015 5th World Congress on Information and Communication Technologies (WICT)*, Dec. 2015, pp. 73–77.

[12] C. Bian, C. Yuan, W. Kuang, and D. Wu, "Evaluation, Classification, and Influential Factors Analysis of Traffic Congestion in Chinese Cities Using the Online Map Data." *Mathematical Problems in Engineering*, pp. 1–10, Sep. 2016.

[13] Y.-y. Chen, Y. Lv, Z. Li, and F. Y. Wang, "Long short-term memory model for traffic congestion prediction with online open data," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2016, pp. 132–137.

[14] H. N. Nguyen, P. Krishnakumari, H. L. Vu, and H. v. Lint, "Traffic COngestion pattern classification using multi-class SVM," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Nov. 2016, pp. 1059–1064.

[15] F. Porikli and X. Li, "Traffic congestion estimation using hmm models without vehicle tracking," in *IEEE Intelligent Vehicles Symposium, 2004*, June 2004, pp. 188–193.

[16] X. Yu, S. Xiong, J. Xiang, J. Mao, M. Liu, and Y. Zhao, "A Campus Traffic Congestion Detecting Method Based on BP Neural Network," in *2015 2nd International Symposium on Dependable Computing and Internet of Things (DCIT)*, Nov. 2015, pp. 93–100.

[17] M. Milojevic and V. Rakocevic, "Distributed road traffic congestion quantification using cooperative vanets," in *2014 13th Annual Mediterranean Ad Hoc Networking Workshop (MED-HOC-NET)*, June 2014, pp. 203–210.

[18] ——, "Short paper: Distributed vehicular traffic congestion detection algorithm for urban environments," in *2013 IEEE Vehicular Networking Conference*, Dec 2013, pp. 182–185.

[19] Transport for London, "Transport for london unified API." [Online]. Available: https://api.tfl.gov.uk/

[20] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, April 2017.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014. [Online]. Available: http://arxiv.org/abs/1409.1556

[22] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard *et al.*, "Tensorflow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI). Savannah, Georgia, USA*, 2016.

[23] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[24] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results," http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html.

[25] D. Braines, A. Thomas, L. Kaplan, M. Sensoy, M. Ivanovska, A. Preece, and F. Cerutti, "Human-in-the-loop situational understanding via subjective bayesian network," in *International Workshop on Graph Structures for Knowledge Representation and Reasoning*. Springer, 2017.

[26] S. Chakraborty, A. Preece, M. Alzantot, T. Xing, D. Braines, and M. B. Srivastava, "Deep learning for situational understanding," in *20th International Conference on Information Fusion*, 2017.

[27] D. Braines, D. Mott, S. Laws, G. de Mel, and T. Pham, "Controlled english to facilitate human/machine analytical processing," *Proc. SPIE*, vol. 8758, pp. 875 808–875 808–13, 2013. [Online]. Available: http://dx.doi.org/10.1117/12.2015907