Detection of Unsigned Ephemeral Road Incidents by Visual Cues

Alex Levering[1], Kourosh Khoshelham[2], Devis Tuia[1], and Martin Tomko[2]

1 **Laboratory of Geo-information Science and Remote Sensing, Wageningen University**
`alex.levering|devis.tuia@wur.nl`

2 **Department of Infrastructure Engineering, University of Melbourne, Australia**
`k.khoshelham|tomkom@unimelb.edu.au`

1 Introduction

Traffic is a highly dynamic environment and ephemeral changes to the on-road conditions impact it continuously. Research in Autonomous Vehicles (AVs) is currently highly focused on dynamic changes, due to the high safety requirements [12]. AVs must be able to detect and respond appropriately to incidents. In a connected traffic data ecosystem [3], AVs will further share traffic information about the state of the road network.

Our research addresses the detection of ephemeral incidents affecting road networks by *first-on-scene vehicles*. An incident is *"... an event, which directly or indirectly can result in considerable reductions or interruptions in the serviceability of a link/route/road network."* [2]. For any unsigned incident (i.e. not yet signposted or tagged incidents such as a recently detected fire), the sensors of the vehicle that is *first on the scene* should detect and assess the danger to the traffic and report it to the connected traffic ecosystem. Efforts have been made to recognize *signed changes* to the environment (i.e., pedestrian and traffic signs) [6, 14], as well as the avoidance of dynamic scenarios (e.g., pedestrians or animals stepping into the traffic [9, 10, 11]). Yet, there has been no research covering the systematic classification and autonomous detection of various types of incidents by first-on-scene vehicles.

In this paper we address the problem of identifying and classifying *unsigned ephemeral on-road incidents* from street-level imagery, such as those acquired by dashcams. We present the approach to the creation of an extensive, labeled street-level image library that supports the detection and classification of on-road incidents using a deep convolutional neural network (CNN). The full results of the experiments will be reported at the workshop.

2 Methodology

Image classification using CNNs has had enormous success [1]. Pre-trained CNNs can now be used for fast image classification from camera frames onboard a vehicle. CNNs require vast amounts of labeled data to be trained. We describe the systematic approach to the collection of a detailed and comprehensive image dataset of on-road incidents, from a taxonomy of incidents through collection from Web sources, incl. data with a partial geographical stratification.

2.1 Taxonomy of Incidents

To label the incident images used by the CNN model, we propose an approach grounded in a systematic taxonomy of incident types, providing a semantic structure for an adaptable set of incident labels (Figure 1). Such adaptability is desirable to for alterations required to

Spatial big data and machine learning in GIScience, Workshop at GIScience 2018.
Editors: Martin Raubal, Shaowen Wang, Mengyu Guo, David Jonietz, Peter Kiefer;
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

reflect changing local deployments (i.e, wildlife, meteorological events). We only consider physical incidents, as non-physical incidents cannot be identified in imagery (e.g., GPS or traffic information system failure). At the top level, we distinguish between signed physical incidents (not tackled here) and unsigned physical incidents. We further distinguish between man-made and natural physical incidents to distinguish the objects and covers altering the status of the road. In total, we consider eight incidents in this initial study: crashes (MAN-O-1), road collapse (MAN-C-1), animal on road (NAT-O-1), treefall (NAT-O-2), snow on road (NAT-C-1), flooding (NAT-C-2), landslides (NAT-C-3), and fires (NAT-C-4).

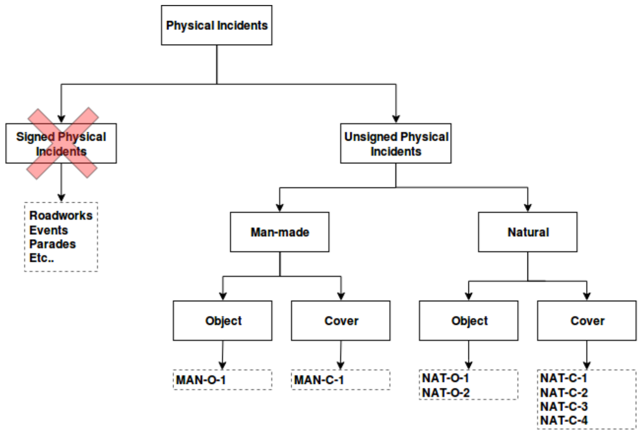

Figure 1 Taxonomy of incidents and their semantic groupings. Each lowest-level indicator (e.g. MAN-O-1) refers to an incident under consideration.

2.2 Data

We collect RGB images captured by image sensors. We opted for optical recognition of incidents as cameras record information that cannot be extracted from LIDAR point clouds (e.g., smoke, fire), and because of the abundance of labeled images and datasets. RGB images are also readily accessible through online search engines. We collect and clean images from four sources: Google Custom Search API [1], Bing Image Search API v7 [2], the Flickr API [3], and the Geograph UK project [4]. We construct queries by grouping synonyms of each topic, such as 'snow on road' and 'street blizzard'. We retain the top-100 images that match queries crafted for each category of taxonomy, as the relevance and quality of images drops noticeably beyond this number. In a subsequent manual cleaning process images are restricted to relevant on-road incidents. We only retain images that have a viewport height of on-road vehicles, comparable to vehicle-mounted cameras. Lastly, we retain only imagery that is *spatially relevant*, i.e., every image relates to a possible incident on the road network. As an example, an animal in itself may not be an incident. It only becomes an incident when it is in a certain spatial context, in this case *close to or on the road*. The detection of such spatial relationships in image classifiers is currently in its infancy [7], and few datasets are built with this intrinsic property in mind – a shortcoming addressed in our collection. In

[1] https://developers.google.com/custom-search/json-api/v1/overview
[2] https://docs.microsoft.com/en-us/rest/api/cognitiveservices/bing-images-api-v7-reference
[3] https://www.flickr.com/services/api/
[4] https://www.geograph.org.uk/

total, we have collected 12.500 images spread across all incident classes. The set contains the following amount of images per class: *Animal on road*: 1321, *Collapse*: 491, *Crash*: 1478, *Fire*: 865, *Flooding*: 2155, *Landslide*: 825, *Snow*: 4744, *Treefall*: 751. The dataset and the query strategy is elaborated upon in more detail during the workshop.

An additional dataset of true-negative images to be used during classification will be gathered from the same sources and from image frames from benchmark datasets such as CityScapes [4]. In total, we aim for a dataset of 40,000 true-negative images to capture a great variety of environments and regular driving conditions. We maintain a 70:20:10 training, validation, and testing split respectively for both true-positive and true-negative subsets.

2.3 CNN Classification

The incident recognition is implemented using CNNs in the PyTorch environment in a multiclass classification task performed on a pre-trained ResNet-34 model [8]. We chose the ResNet architecture to leverage skip connections to reduce overfitting, as well as its state-of-the-art performance and ease of training. We use a model pre-trained on the ImageNet dataset [5] which is then re-trained across all layers. We don't consider multi-label cases (e.g. a fire and a car crash visible in the same image). During training we track the loss and classification accuracy, retaining the model with the lowest validation loss as the best model. The final overall accuracy is given by the F1-score and the overall accuracy. A qualitative analysis of visual triggers will be supplied through the use of *Gradient-weighted Class Activation Mapping* [13], which will help to highlight the model's visual cues for each class. Lastly, we also report the accuracy of an experiment using a geographical stratification for the classes *Animal on road*, *flooding*, and *Snow*, which contain geotagged images from the Geograph dataset.

3 Discussion and Conclusions

Matches for images tagged with *incident* – a somewhat jargon term for events on roads – are rare. Our searches therefore combine terms that can either be identified in the images, or in the text nearby. The applied restrictions limit the amount of relevant images, leading to problems with sourcing examples for certain classes. This may noticeably limit the amount of possible representations of incidents covered by the final dataset. Currently, we are training our CNN model, varying all relevant hyperparameters until satisfactory convergence. The final model performance will be reported at the workshop using the F1-score and the top-1 accuracy. Lastly, we will report the geographically stratified accuracy of the three classes (*animals, flooding, snow*) present in the Geograph dataset.

We anticipate problems due to classification correlation for themes present in certain images (e.g., landslides may correlate with rocky cliff-faces). Results for all classes may thus not be equally robust. Possible continuations of this research may focus on more efficient use of image samples from sparsely represented classes to increase their representational power. We envisage to publish the subset of images licensable under a CC-by license along with the pre-trained network. This will allow for further experimentation and improvements on the benchmark for this task that has so far not been broadly covered in the literature and where significant progress is possible.

References

1 Rodrigo Benenson. Classification datasets results, 2016. URL: `http://rodrigob.github.io/are_we_there_yet/build/classification_datasets_results.html#43494641522d3130`.

2 Katja Berdica. An introduction to road vulnerability: what has been done, is done and should be done. *Transport Policy*, 9(2):117–127, April 2002. URL: `http://www.sciencedirect.com/science/article/pii/S0967070X02000112`, `doi:10.1016/S0967-070X(02)00011-2`.

3 Riccardo Coppola and Maurizio Morisio. Connected Car: Technologies, Issues, Future Trends. *ACM Comput. Surv.*, 49(3):46:1–46:36, October 2016. URL: `http://doi.acm.org/10.1145/2971482`, `doi:10.1145/2971482`.

4 Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes Dataset for Semantic Urban Scene Understanding. *arXiv:1604.01685 [cs]*, April 2016. arXiv: 1604.01685. URL: `http://arxiv.org/abs/1604.01685`.

5 Jia Deng, Wei Dong, Richard Socher, Li-jia Li, Kai Li, and Li Fei-fei. Imagenet: A large-scale hierarchical image database. In *In CVPR*, 2009.

6 Zoltán Fazekas, Gábor Balázs, László Gerencsér, and Péter Gáspár. Locating road-works sites via detecting change in lateral positions of traffic signs measured relative to the ego-car. *Transportation Research Procedia*, 27:341–348, January 2017. URL: `http://www.sciencedirect.com/science/article/pii/S2352146517309018`, `doi:10.1016/j.trpro.2017.12.004`.

7 Mandar Haldekar, Ashwinkumar Ganesan, and Tim Oates. Identifying Spatial Relations in Images using Convolutional Neural Networks. *arXiv:1706.04215 [cs]*, June 2017. arXiv: 1706.04215. URL: `http://arxiv.org/abs/1706.04215`.

8 Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *arXiv:1512.03385 [cs]*, December 2015. arXiv: 1512.03385. URL: `http://arxiv.org/abs/1512.03385`.

9 M. Jeong, B. C. Ko, and J. Y. Nam. Early Detection of Sudden Pedestrian Crossing for Safe Driving During Summer Nights. *IEEE Transactions on Circuits and Systems for Video Technology*, 27(6):1368–1380, June 2017. `doi:10.1109/TCSVT.2016.2539684`.

10 B. Pan and H. Wu. Urban traffic incident detection with mobile sensors based on SVM. In *2017 XXXIInd General Assembly and Scientific Symposium of the International Union of Radio Science (URSI GASS)*, pages 1–4, August 2017. `doi:10.23919/URSIGASS.2017.8104994`.

11 Khaled Saleh, Mohammed Hossny, and Saeid Nahavandi. Kangaroo vehicle collision detection using deep semantic segmentation convolutional neural network. *DICTA 2016 : Proceedings of the IEEE International Conference on Digital Image Computing: Techniques and Applications*, January 2016. URL: `http://dro.deakin.edu.au/view/DU:30092160`, `doi:10.1109/DICTA.2016.7797057`.

12 Brandon Schoettle and Michael Sivak. A survey of public opinion about autonomous and self-driving vehicles in the US, the UK, and Australia. 2014.

13 Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *arXiv:1610.02391 [cs]*, October 2016. arXiv: 1610.02391. URL: `http://arxiv.org/abs/1610.02391`.

14 H. Yong and X. Jianru. Real-time traffic cone detection for autonomous vehicle. In *2015 34th Chinese Control Conference (CCC)*, pages 3718–3722, July 2015. `doi:10.1109/ChiCC.2015.7260215`.