‘Innovation and the Application of Knowledge for More Effective Policing’

N8 Policing Research Partnership Catalyst Project

PROJECT TITLE: Emerging Technology and Big Data Analytics: Realising the Potential of Automatic Number Plate Recognition

INTRODUCTION

In recent years, the replacement of vehicle number plates to avoid detection has become a major policing issue. We refer here to those from any other vehicle as false plates, and to those from identical make, model and colour vehicles as clone plates. The UK’s Automatic Number Plate Recognition (ANPR) camera network offers the potential to explore the pervasiveness of this problem. This project sought to explore means to identify false and clone number plates from ANPR images using machine learning methods, and to thereby help realise the potential of police investment in ANPR infrastructure.

KEY FINDINGS

- Machine learning proved successful at identifying false plates from ANPR images. It holds significant potential as a policing tool. Validation indicates our analytical workflow is 92% accurate in identifying false number plates. The tool that we developed is termed the False Plate Classifier (FPC).
    - In an unseen sample of 130,000 images, our analytics estimate the proportion of false plates where the vehicle make, model or colour differs to that registered, to be in the order of 0.03% (1 in 3700 vehicles).
- The FPC provides a proof-of-concept detection algorithm which, in its current format, is highly accurate in detecting false plates on vehicles with a different make, model or colour to that registered. Further model development and data sources are required to progress the detection of clone plates, but our results are highly encouraging.
    - We note also that the transfer of ANPR data from police to a university, in the context of GDPR, requires considerable commitment and technical skill from both parties but, as demonstrated here, is feasible. Data cleaning, handling and wrangling issues are complex, and require considerable investment. Machine learning requires specialist research expertise.
- The development of a software tool for police use, and an operational trial in a police control room, would be the appropriate next steps alongside continued development of detection methods.
- Further investment in ANPR-related research, and continued development of a False Plate Classifier, would help realise the potential of the significant and continuing investment in ANPR infrastructure.

THE RESEARCH TEAM

The research involved teams from West Yorkshire Police (WYP) and the University of Leeds (UoL) with input from the University of Leicester. The WYP team was led by Supt Mark Jessop (WYP ANPR lead), Neil Hudson (Force ANPR Manager), Pip Reed (ANPR Analyst), plus technical, legal and operational staff. The UoL team was led by Prof Graham Farrell, Dr Daniel Birks, Prof Nick Malleson, and Dr Eva Heinen, with advice from Dr Andrew Newton (Leics), plus technical and legal support at UoL. The machine learning development was undertaken by Dr Adam Hardy who was employed as Research Fellow at UoL.

BACKGROUND

The project arose in response to a call by West Yorkshire Police for police-university collaboration on ANPR, to bring university expertise together with police operational experience and data. This project responded to that call. It sought to catalyse a process by which value might be added to the role of ANPR in tackling crime by mobile offenders, through the improved detection of false licence plates.

PROJECT SUMMARY

While ANPR has existed for two decades, there has been major investment in police ANPR infrastructure in recent years. Millions of ANPR images are captured every day across the UK, with often over 2M in West Yorkshire alone. This makes for a major 'big data' source. The scale of the data dictate that analysis of patterns, trends and other aspects can be resource intensive and may require highly specialised analytical skills. Consequently, some operational aspects of policing can be made more difficult because the needle (of offending) is hidden in an overwhelmingly large haystack (of images). In response, this project combined police ANPR expertise and data with university crime science and data analytic expertise.

The transfer of ANPR data from police to a university, in the context of GDPR, required considerable commitment, technical and legal skill from both parties. This is an important consideration for further research in this area. The project achieved the transfer of around 5 million ANPR image records from 16 motorway ANPR cameras. These cameras were selected because their images only showed the lower front part of vehicles and so did not capture images of windscreens or persons. This issue, which relates to GDPR, should be considered by further research in this area. Further research should also consider issues relating to the representativeness of the sample.

The database of 5 million images was small relative to the total – equivalent to only two or three days' images for the force as a whole. We soon realised that, for our development purposes, a subsample of the data would allow greater (computationally faster) flexibility. This is a useful and practical consideration for further development research in this area.

The project employed Dr Adam Hardy as Research Fellow. Dr Hardy's expertise in machine learning underpinned the project management's decision to pursue this core analytic aspect of the work as that which would most likely produce fruitful and significant results within the timeframe. That focus enabled us, we believe, to produce the first methodologically justifiable automated approach to false number plate detection.

METHODOLOGY

The machine learning component of the research involved the development of a convolutional neural network (CNN). The technicalities of the work are complex but the concept is straightforward, and this report draws heavily on the methodological report prepared by the Research Fellow (Hardy, A. 2019. '*Machine Learning for the Identification of Cloned Number plates*', School of Law, University of Leeds). The CNN examined the ANPR images to identify the vehicle's make, model and colour. These characteristics were then compared to the recorded make model and colour on the DVSA database of registered vehicles. Where discrepancies existed, vehicles were potentially using false or clone number plates. Some of the discrepancies were found to be due to other issues: including where the ANPR software incorrectly read the number plate, or where the quality of the image was poor.

In the final validation sample of 130,000 images, the proportion of number plates classified as false was 0.027% or around 1 in 3700 vehicles. While this may seem to be a low rate, it is one of the first methodologically justifiable estimates of the extent of the problem. It may also suggest that the majority of the problem involves the use of vehicles where offenders purposively choose number plates from identical make, model and colour vehicles – cloned number plates. While our initial focus in this exploratory project was to develop detection methods for false number plates, this research demonstrates an analytical framework that, given additional vehicle registration data, may permit the detection of more subtle difference between vehicles, such as vehicle sub-models.

Further background details to methodology:

Prior to the arrival of the WYP image data at UoL, preparatory research was undertaken on the Carvana and ComCars datasets, which are publicly available. The DVSA online MOT database was scraped to obtain vehicle-level information that could then be cross-referenced to the WYP ANPR images to identify the vehicle make, model, and colour. Cleaning of the ANPR and DVSA data included aspects such as ensuring between-database compatibility in vehicle names (e.g. MERCEDES versus MERCEDES-BENZ). Identification of suspect vehicles was achieved by analysing the ANPR image of the car and comparing vehicle characteristics classified by machine learning algorithms to those registered in the DVSA data (which served as the publicly available proxy for the DVLA database). A convolutional neural network (CNN) was deployed on the image to determine the probability of it being each known colour, make and model. The probability of it being the colour, make and model corresponding to its number plate was then extracted, and a decision tree built on these probabilities to identify suspect vehicles. After deploying the algorithm on a previously unseen testing subset of 130,000 images, 1237 images were identified as potential clones. Of these, 1006 were found to be mis-classifications and 90 were bad images (such as blurred black and white images that needed to be removed). In addition, 1006 of the images were false plate reads (where the ANPR software read the plate incorrectly), which appeared as suspect vehicles as far as the algorithm was concerned. This left 35 images of plates which were likely to be of genuine interest to the police.

CONCLUSION

Police use of ANPR seeks to reduce crime via mechanisms of detection and deterrence. Research of its effectiveness is scarce, but it is likely to have reduced driving offences and criminal use of vehicles. However, some offenders have sought to circumvent ANPR (both public and private) through the use of false or clone number plates.

This project provides a proof-of-concept of the potential to apply machine learning for the detection of false or cloned licence plates. The preliminary finding is that vehicles with different make, model or colour to the registered vehicle comprise 0.03% (around 1 in 3700) of vehicles in our sample. This is a lower estimate than anticipated and suggests that offenders make the extra effort to match vehicle model and colour.

The progress made here suggests a need for progress towards operational use in two ways. First, we recommend continued refinement of a False Plate Classifier using machine learning approaches to include additional distinguishing features to detect clone plates on vehicles of the same model and colour as that registered. This approach may combine multiple techniques and include feature-detection algorithms that compare specific details of the number plate across multiple images, and a secondary machine learning method for classifying visible vehicle features such as trim. Second, we recommend the parallel development of a software tool for police use, enabling operational trials to be conducted in a police control room.

IMPLICATIONS FOR FURTHER RESEARCH

The project provided proof-of-concept of the potential for machine learning to be developed to identify suspect vehicle number plates. A series of practical lessons for further research were identified relating to data transfer, data cleaning, handling and wrangling, the development of machine learning algorithms, and necessary requirements for further investigation of cloned plates. Here, a number of avenues of further research are suggested.

First, automated techniques for identifying number plate misreads could be developed. While our algorithms were effective at detecting misreads they were currently classified as suspect vehicles (necessarily). One analytic option would be to search for commonly misread letters such as M and N and utilise this information to flag potential mis-reads over actual suspect vehicles. As ANPR technology advances this is likely to become less of a problem.

Further research is needed to both advance existing and explore new methods by which clones can be identified. While the present approach was capable of detecting vehicles which differed from those registered by make, model or colour, the lack of more detailed vehicle registration data was a constraint. With relevant sub-model data (i.e. VW Golf S vs VW Golf GTI), the method of image classification proposed here could likely be adapted to include sub-model. Furthermore, other approaches may be possible: one insightful method suggested by WYP is the possibility of looking for differences in individual number plates. Cross-checking multiple images of the same vehicle could allow automated scrutiny of differences in number plates. It is also likely that data from the False Plate Classifier could be combined with other data (space, time, incident etc.) to support a more holistic understanding of criminal use of false and clone number plates.

Authors: Graham Farrell, Adam Hardy & Dan Birks (University of Leeds) and West Yorkshire Police – **December 2019**

Further information – Graham Farrell (G.Farrell@leeds.ac.uk)