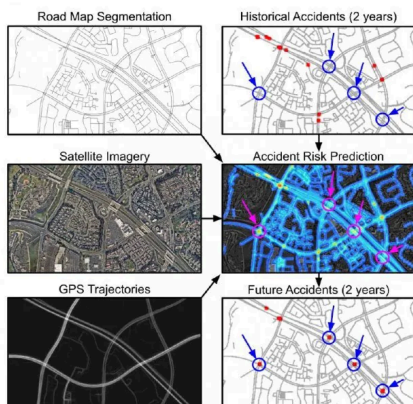


AI Predicts Accident Hot-Spots From Satellite Imagery and GPS Data

Originally published at <https://www.unite.ai/ai-predicts-accident-hot-spots-from-satellite-imagery-and-gps-data/>



Researchers from MIT and the Qatar Center for Artificial Intelligence have developed a machine learning system that analyzes high-resolution satellite imagery, GPS coordinates and historical crash data in order to map potential accident-prone sections in road networks, successfully predicting accident 'hot spots' where no other data or previous methods would indicate them.



Middle right, predictive accident hot-spots emerge from collating three sources of data. Areas highlighted in circles are 'high risk' predictions that actually have no historical accident history. Source:

https://openaccess.thecvf.com/content/ICCV2021/papers/He_Inferring_High-Resolution_Traffic_Accident_Risk_Maps_Based_on_Satellite_Imagery_ICCV_2021_paper.pdf

The system offers bold predictions for areas in a road network that are likely to become accident black-spots, even where those areas have zero history of accidents. Testing the system over data covering four years, the researchers found that their predictions for these 'no history' potential accident hazard zones were borne out by events in subsequent years.

The [new paper](#) is called *Inferring high-resolution traffic accident risk maps based on satellite imagery and GPS trajectories*. The authors predict uses for the new architecture beyond accident prediction, hypothesizing that it could be applied to 911 emergency risk maps or systems to predict the likelihood for demand for taxis and ride-share providers.

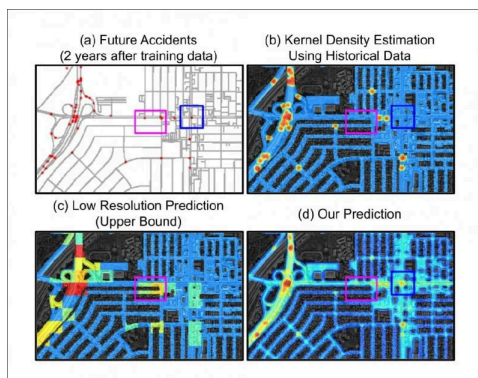
Prior similar efforts have attempted to create similar incident-predictors from low-resolution maps with high bias, or else to leverage accident frequency as a key, which led to high-variance, inaccurate predictions. Instead, the new project, which covers four major US cities totaling 7,488 square kilometers, outperforms these earlier schemes by collating more diverse forms of data.

Sparse Data

The problem the researchers face is sparse data – very high volumes of accidents will inevitably be noticed and addressed without the need for machine analytics, but more subtly dangerous correlations are difficult to identify.

Previous accident prediction systems center on [Monte Carlo estimation](#) of historical accident data, and can provide no effective prediction mechanism where this data is lacking. Therefore the new research studies road network sections with similar traffic patterns, similar visual appearance and similar structure, inferring a disposition to accidents based on these characteristics.

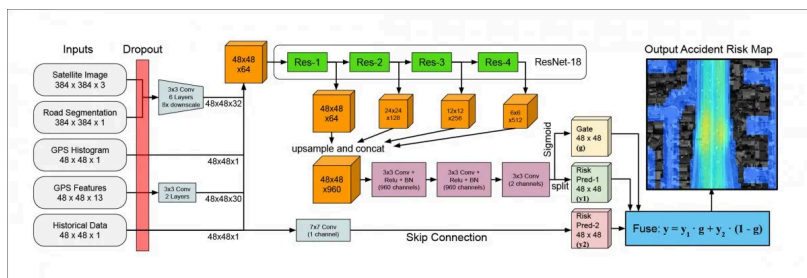
It's a 'shot in the dark' that seems to have unearthed fundamental accident indicators, which could be utilized in the design of new road networks.



Kernel Density Estimation (KDE) has been used to highlight historical traffic accident hot-spots, failing to predict future accident locations. In the upper left image we see where KDE has predicted accidents in the blue box region, versus where the accidents generally localized (adjacent). Bottom right, a comparison of KDE prediction failure to the accurate prediction (blue box) of the MIT system.

The authors note that GPS trajectory data offers information on the flow, speed and density of traffic, while satellite imagery of the area adds information about lane disposition, and the number of lanes, as well as the existence of a hard shoulder and the presence of pedestrians.

Contributing author Amin Sadeghi, from Qatar Computing Research Institute (QCRI) [commented](#) "Our model can generalize from one city to another by combining multiple clues from seemingly unrelated data sources. This is a step toward general AI, because our model can predict crash maps in uncharted territories." and continued "The model can be used to infer a useful crash map even in the absence of historical crash data, which could translate to positive use for city planning and policymaking by comparing imaginary scenarios".



The architecture of the traffic prediction system generates an accident risk map at a 5-meter resolution, which the authors state is critical to distinguish different residential roads.

The project was evaluated on crashes and lateral data covering a period between 2017-18. Predictions were then made for 2019 and 2020, with several 'high risk' locations emerging even in the absence of any historical data that would normally predict this.

Achieving Useful Generalization

[Overfitting](#) is a critical risk in a system fueled by sparse data, even where, as in this case, there are two additional sources of supporting data. Where an incidence is low, excessive assumptions can be drawn from too few examples, leading to an algorithm that is expecting a very particular, narrow band of possible circumstances, and which will fail to identify broader probabilities.

Therefore, in training the model the researchers randomly 'dropped out' each input source as a 20% probability, so that areas with less (or no) accident data can be considered as the model trains towards generalization, and so that parallel data sources can act as a representative proxy for missing information for any particular study of an intersection or section of road.

Evaluation

The model was tested on a dataset comprising nearly 7,500km of urban area in Boston, Los Angeles, Chicago and NYC. The dataset was organized in the form of 1,872 2kmx2km tiles, each containing satellite imagery from [MapBox](#), with road segmentation masked via data from OpenStreetMap. Both the base imagery and the segmentation maps have a resolution of 0.625 meters.

The GPS data comes in the form of a proprietary dataset collected between 2015-17 over the four cities, totaling 7.6 million kilometers of GPS trajectories at a 1-second sampling rate.

The project also exploits 4.2 million records covering 2016-2020 in the [US Accidents Dataset](#). Each record includes timestamps and other metadata.

The first two years of historical data were fed to the model, and the final two years used for training and evaluation, enabling the researchers to establish the accuracy of the system over two years in a short time-frame.

The system was tested with and without historical data, and was found to successfully capture the underlying risk distribution across all cases, notably improving on prior KDE-based methods (see above).

Roads Forward

The authors contend that their system can be applied to other countries with little architectural modification, even in locations where accident data is not available. Additionally, the authors propose their research as a possible adjunct to city planning design for new urban developments.

Lead author Songtao He commented on the new work:

"By capturing the underlying risk distribution that determines the probability of future crashes at all places, and without any historical data, we can find safer routes, enable auto insurance companies to provide customized insurance plans based on driving trajectories of customers, help city planners design safer roads, and even predict future crashes."

Though the paper indicates that the code for the system has been released on GitHub, the link to the code is not active, can't currently be found by a search, and presumably will be included in a later revision.

The research has potential to be incorporated into popular consumer-level GPS-based traffic apps and route planners, according to Songtao He:

"If people can use the risk map to identify potentially high-risk road segments, they can take action in advance to reduce the risk of trips they take. Apps like Waze and Apple Maps have incident feature tools, but we're trying to get ahead of the crashes — before they happen,"

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More