Inferring high-resolution traffic accident risk maps based on satellite imagery and GPS trajectories

Songtao He, Mohammad Amin Sadeghi, Sanjay Chawla, Mohammad Alizadeh, Hari Balakrishnan, Samuel Madden

Massachusetts Institute of Technology

CSAIL

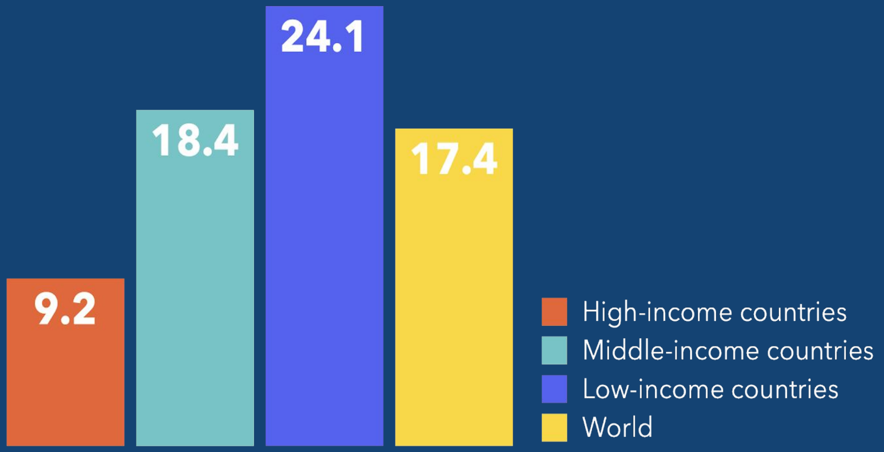
معهد قطر لبحوث الحوسبة
Qatar Computing Research Institute
Member of Qatar Foundation عضو في مؤسسة قطر

Rising Crash Rates and Costs

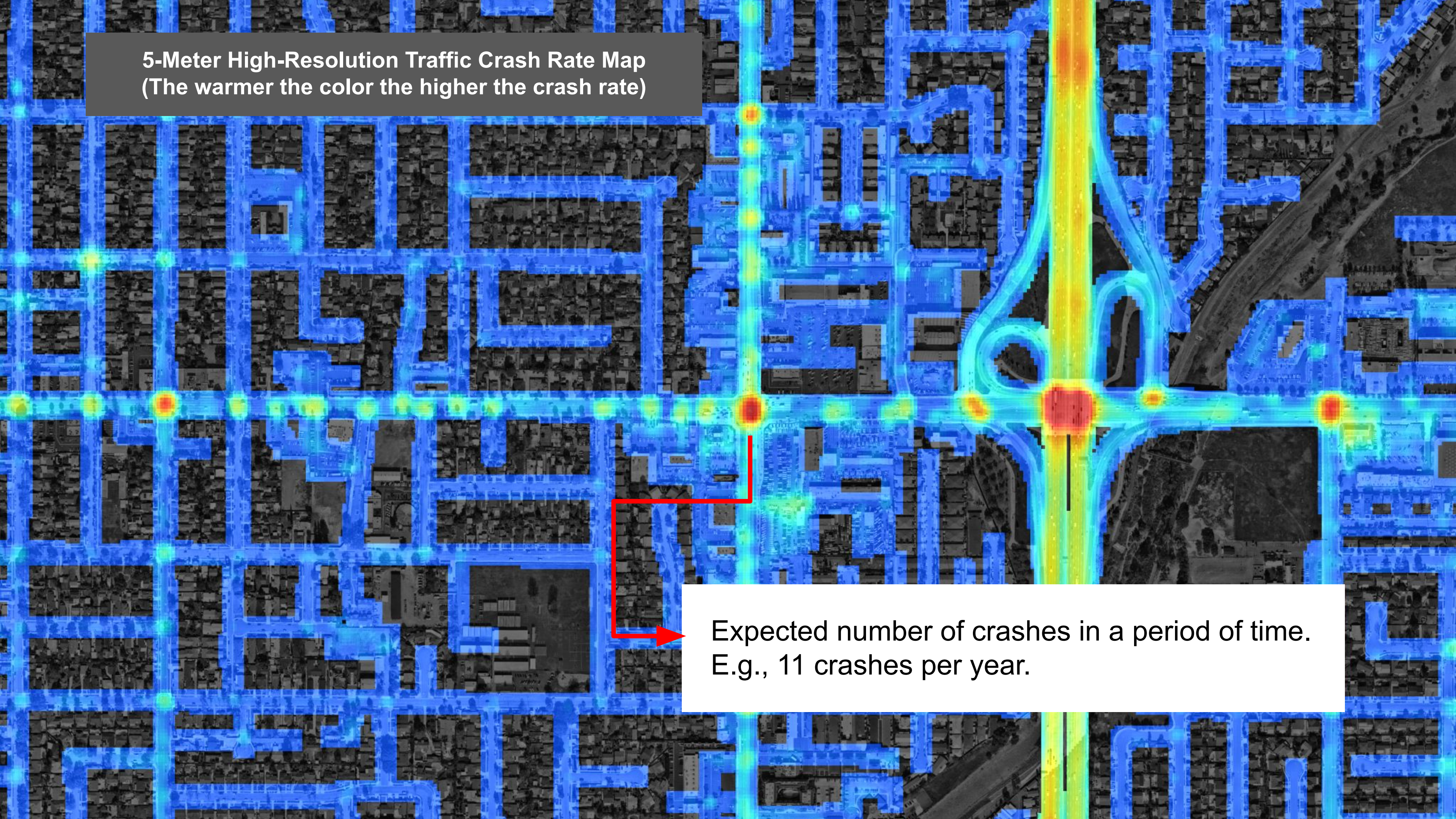
50 MILLION
Road injuries per year

$1.8 TRILLION
Loss costs per year

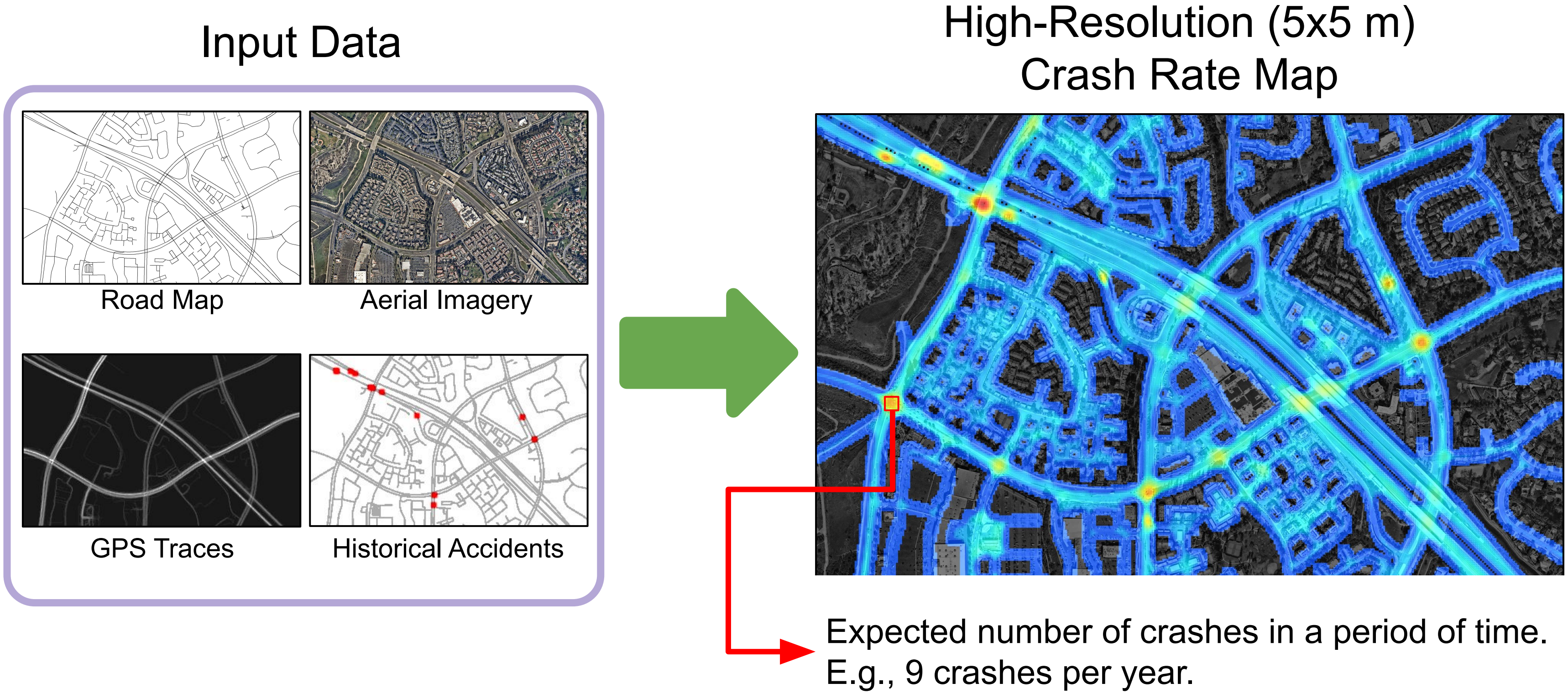
1.35 MILLION
Road deaths per year
(50% of COVID)



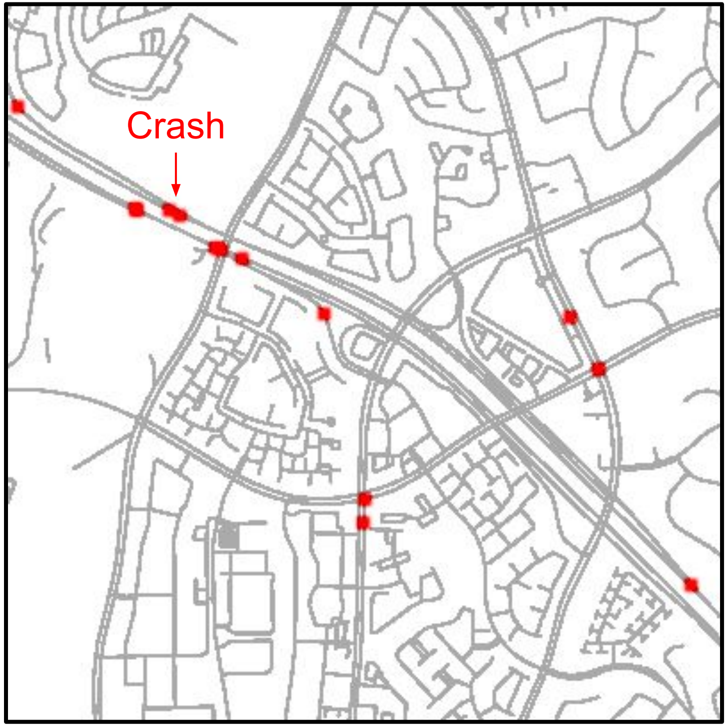
Deaths per 100K inhabitants per year


5-Meter High-Resolution Traffic Crash Rate Map
(The warmer the color the higher the crash rate)
Expected number of crashes in a period of time.
E.g., 11 crashes per year.

Crash Rate Maps

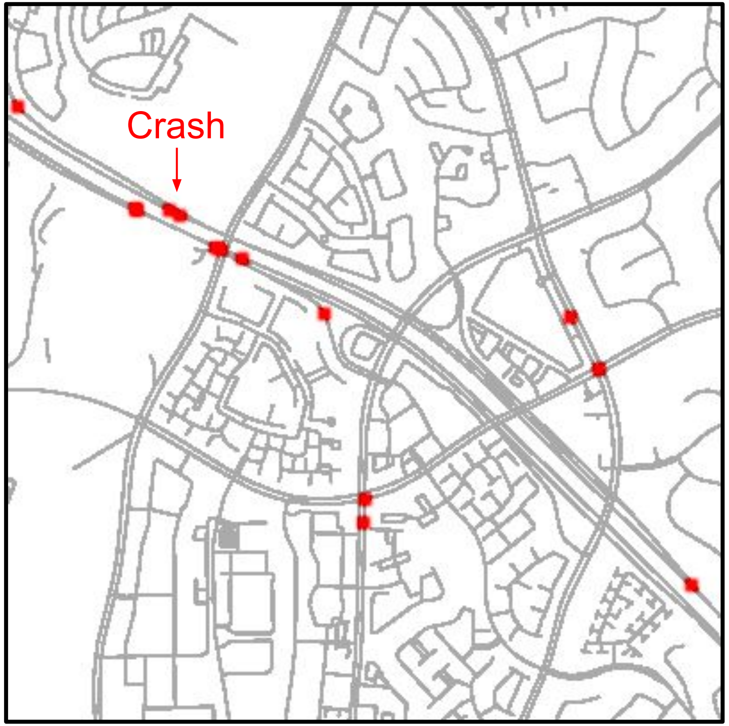


Challenge: Crash Data is Very Sparse

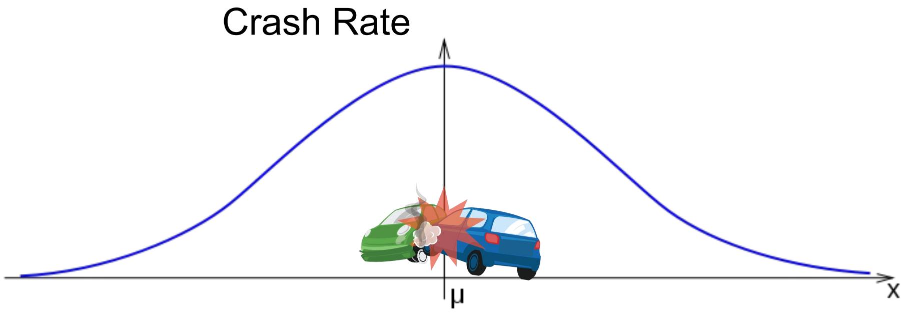


Crashes in 2017 and 2018
at certain locations in Los Angeles

Traditional: Kernel Density Estimation on Historical Data



Crashes in 2017 and 2018
at certain locations in Los Angeles



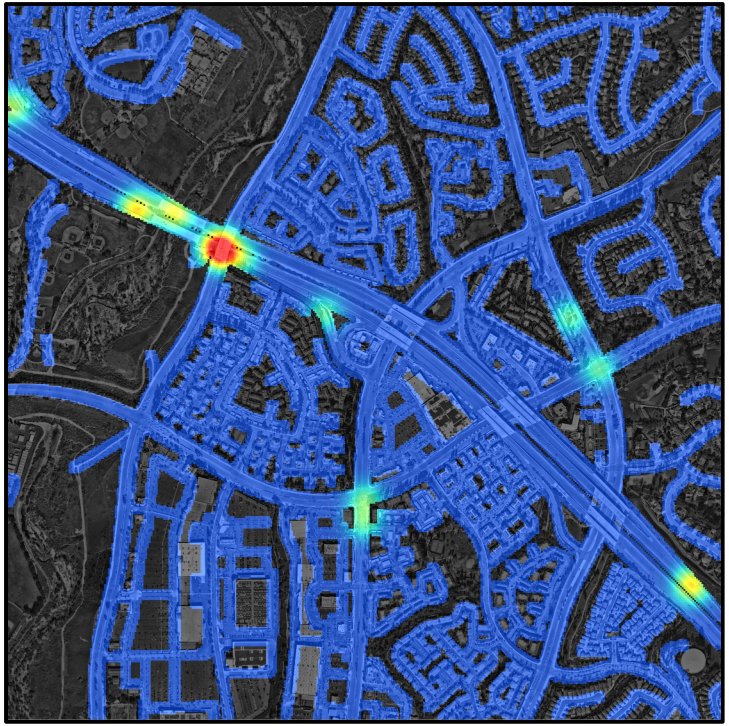
Kernel Density Estimation on Historical Data

[24] L. T. Truong and S. V. Somenahalli. Using gis to identify pedestrian-vehicle crash hot spots and unsafe bus stops. *Journal of Public Transportation*, 14(1):6, 2011. 2

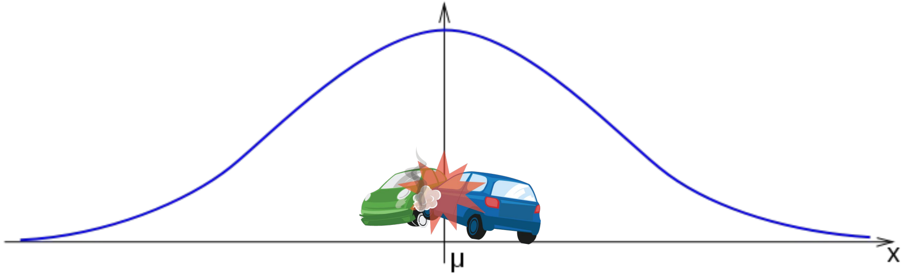
[25] Z. Xie and J. Yan. Kernel density estimation of traffic accidents in a network space. *Computers, environment and urban systems*, 32(5):396–406, 2008. 2

[26] Z. Xie and J. Yan. Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of transport geography*, 31:64–71, 2013. 2

Kernel Density Estimation on Historical Data

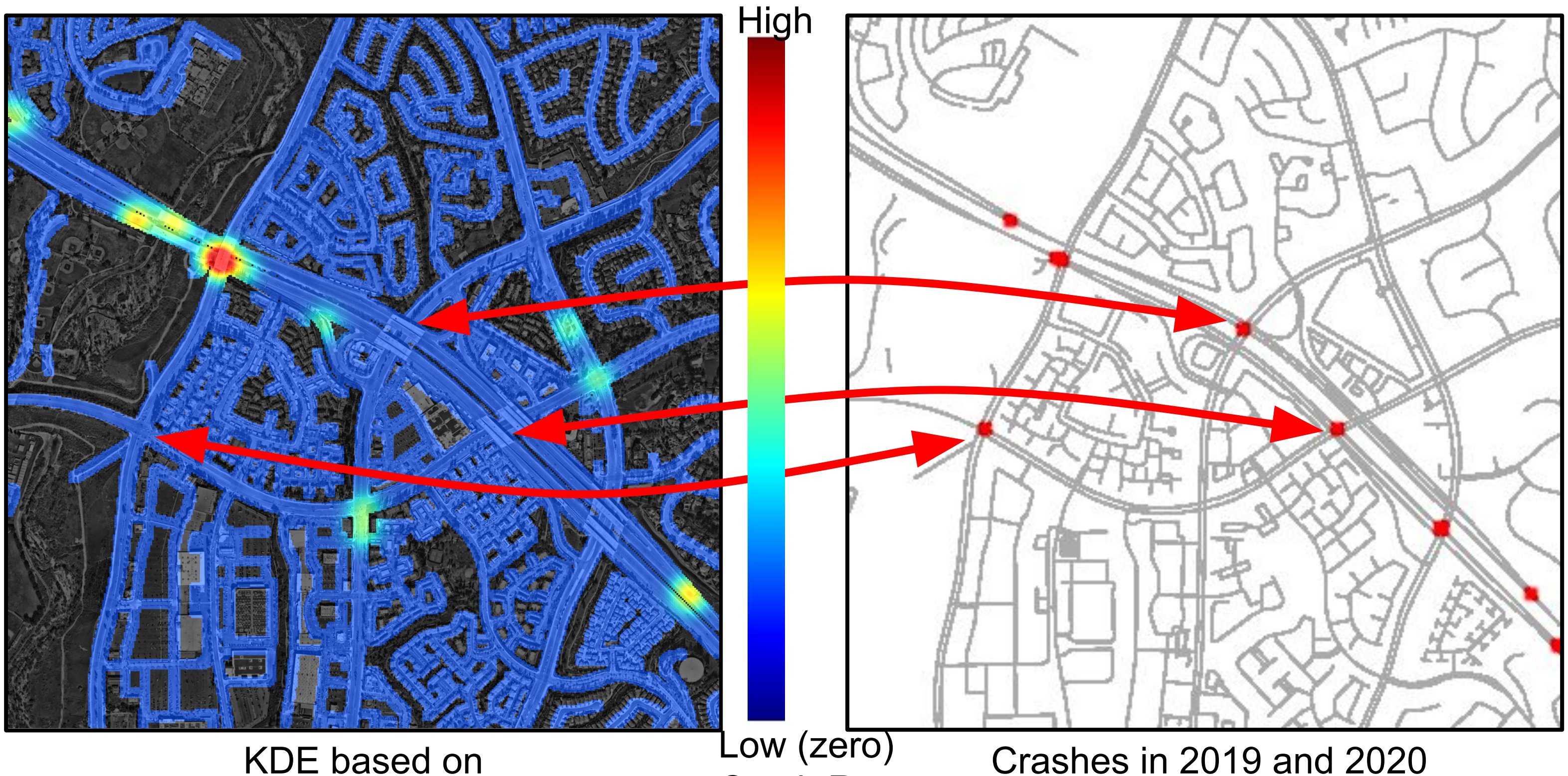


KDE based on
Crashes in 2017 and 2018



Kernel Density Estimation on Historical Data

Kernel Density Estimation on Historical Data



KDE based on
Crashes in 2017 and 2018

Crashes in 2019 and 2020

Kernel Density Estimation on Historical Data

Crashes in 2017 and 2018

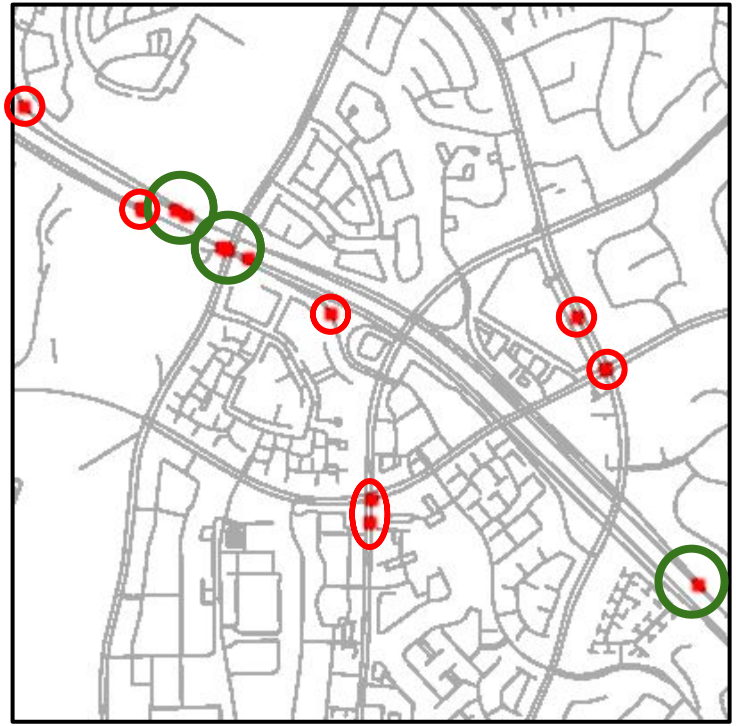
Crashes in 2019 and 2020

Kernel Density Estimation on Historical Data

Crashes in 2017 and 2018

Crashes in 2019 and 2020

Kernel Density Estimation on Historical Data



Crashes in 2017 and 2018

Crashes in 2019 and 2020

Learning-Based Solution

Crash
Rate Prediction

Input data at a location,
e.g., road layout and
aerial imagery.

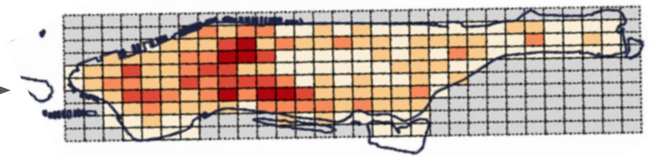
$$y = f(x)$$

Model
e.g., a neural network model

Prior Work

Year	Authors	Resolution	Method	Input data
2005	Chang et al. [9]	Entire highway	Decision Tree	Road map, average daily traffic (AADT), weather
2005	Chang et al. [8]	Entire highway	Neural Networks	Road map, average daily traffic (AADT), weather
2007	Caliendo et al. [7]	Entire highway	Max. Likelihood	Road map, AADT, slope and presence of junctions
2016	Chen et al. [11]	$500\text{m} \times 500\text{m}$	SdAE [4]	GPS trajectories, historical accidents
2017	Yuan et al. [29]	road segments	Deep networks	Historical Accidents, road map, weather
2017	Najjar et al. [17]	$150\text{m} \times 150\text{m}$	Pre-trained Alex-net	Satellite imagery, accident history
2018	Ren et al. [22]	$1\text{km} \times 1\text{km}$	LSTM	Historical accidents
2018	Chen et al. [10]	$500\text{m} \times 500\text{m}$	SdAE [4]	Traffic flow (from plate recognition system), accident history
2018	Yuan et al. [28]	$5\text{km} \times 5\text{km}$	ConvLSTM	Traffic volume, road condition, weather, satellite imagery
2019	Bao et al. [3]	$> 360\text{m}$	STCL-Net	Crash, GPS, road, land use, population and weather data
2020	Zhou et al. [31]	$1.5\text{km} \times 1.5\text{km}$	RiskSeq	Traffic flow, road network, weather and accident history

Table 1 in our ICCV paper*



*Inferring high-resolution traffic accident risk maps based on satellite imagery and GPS trajectories (ICCV 2021)
Songtao He, Mohammad Amin Sadeghi, Sanjay Chawla, Mohammad Alizadeh, Hari Balakrishnan, Samuel Madden

Learning-Based Solution

Crash
Rate Prediction

Input data at a location,
e.g., road layout and
aerial imagery.

$$y = f(x)$$

Model
e.g., a neural network model

Training Dataset

Input data ***x***	Target crash rate ***y***
x_1	y_1
x_2	y_2
...	...
x_N	y_N

Learning-Based Solution

Crash
Rate Prediction

Input data at a location,
e.g., road layout and
aerial imagery.

$$y = f(x)$$

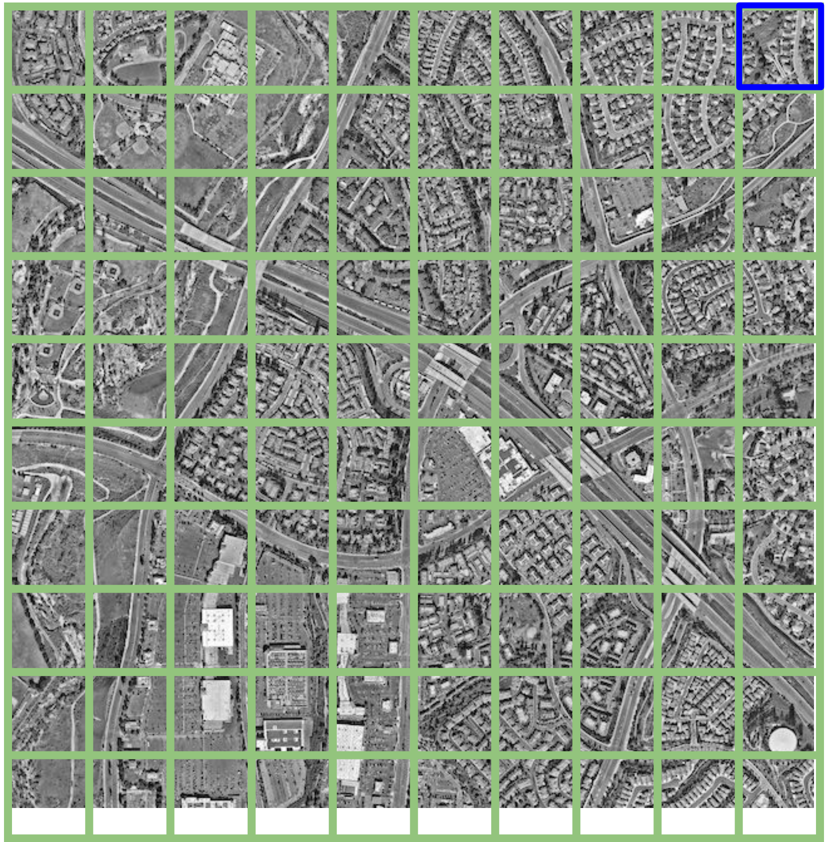
Model
e.g., a neural network model

Training Dataset

Input data ***x***	Target crash rate ***y***
x_1	y_1
x_2	y_2
...	...
x_N	y_N

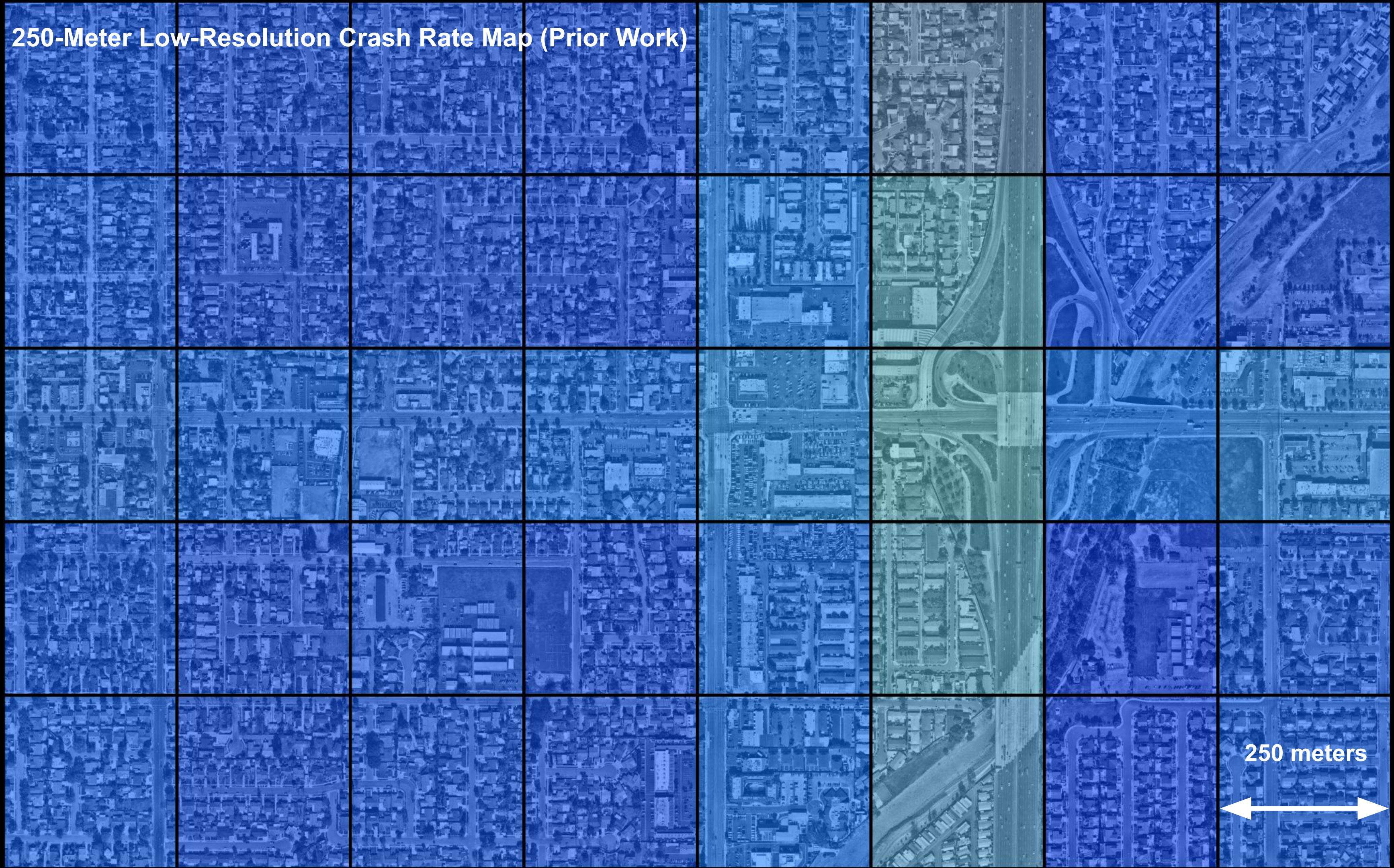
High resolution → Small ***y*** value → Hard to obtain

Target crash rates are hard to obtain at high resolution

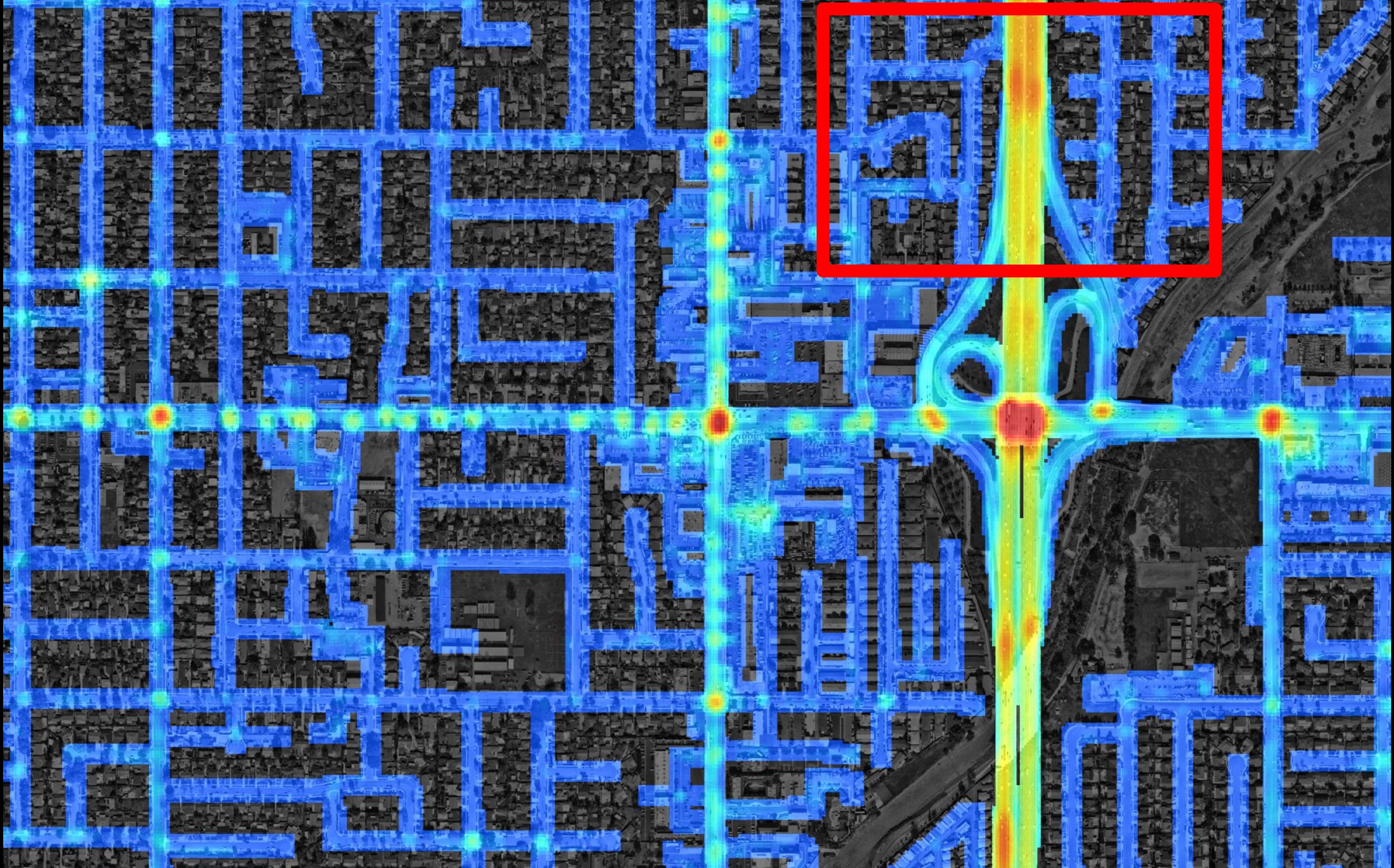


If annual rate is only 0.1 (typical), how do we estimate it?

A grid cell

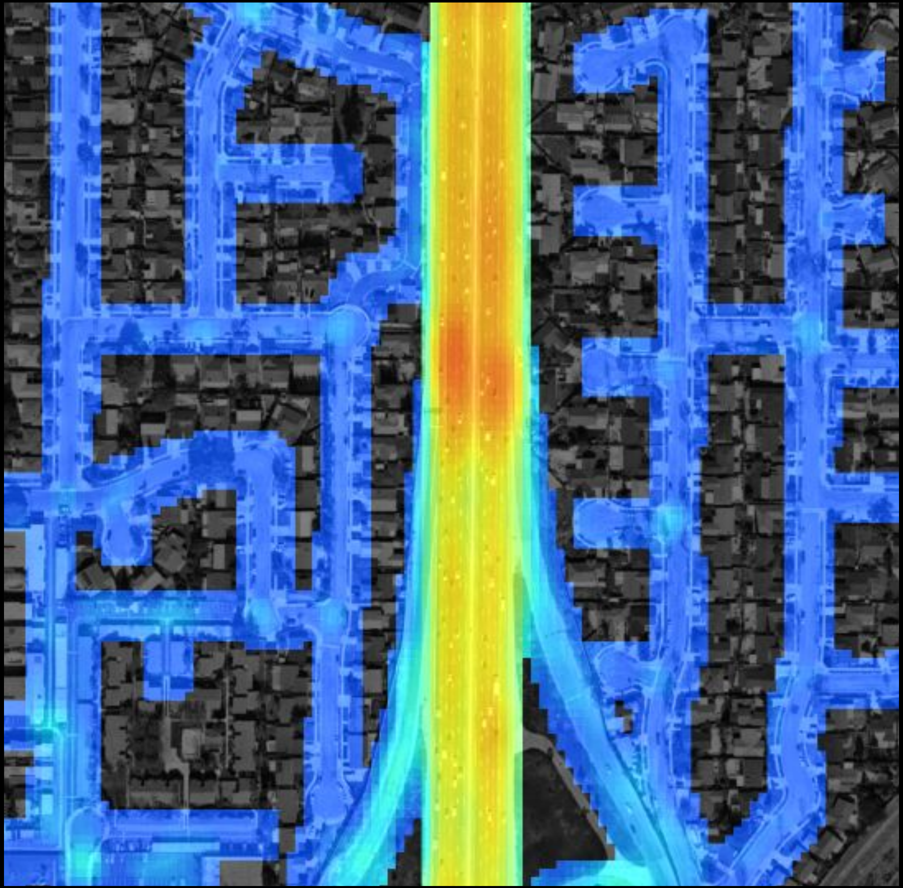

250-Meter Low-Resolution Crash Rate Map (Prior Work)
250 meters


5-meter
resolution



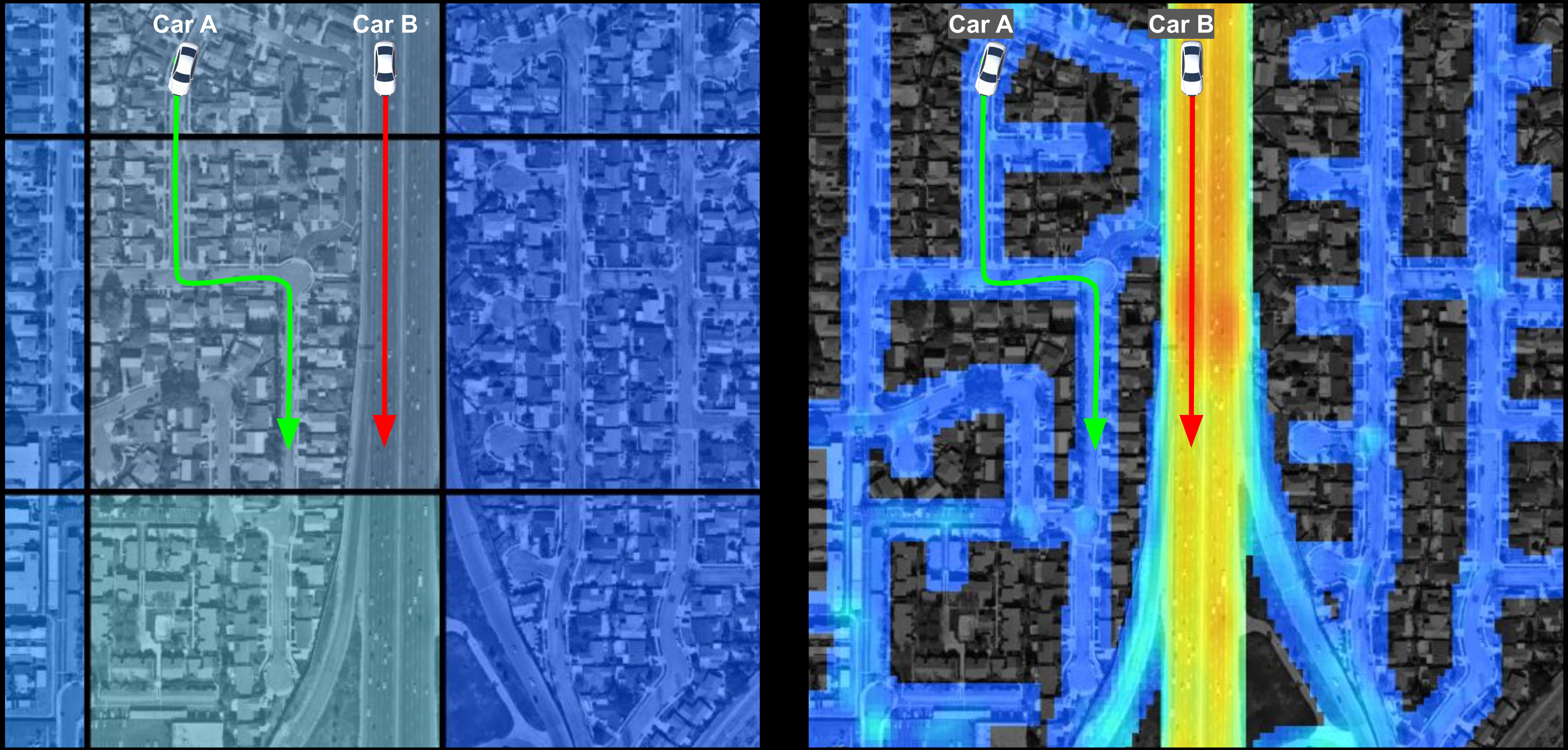
250-Meter Low-Resolution Crash Rate Map (Prior Work)

5-Meter High-Resolution Crash Rate Map (This Work)



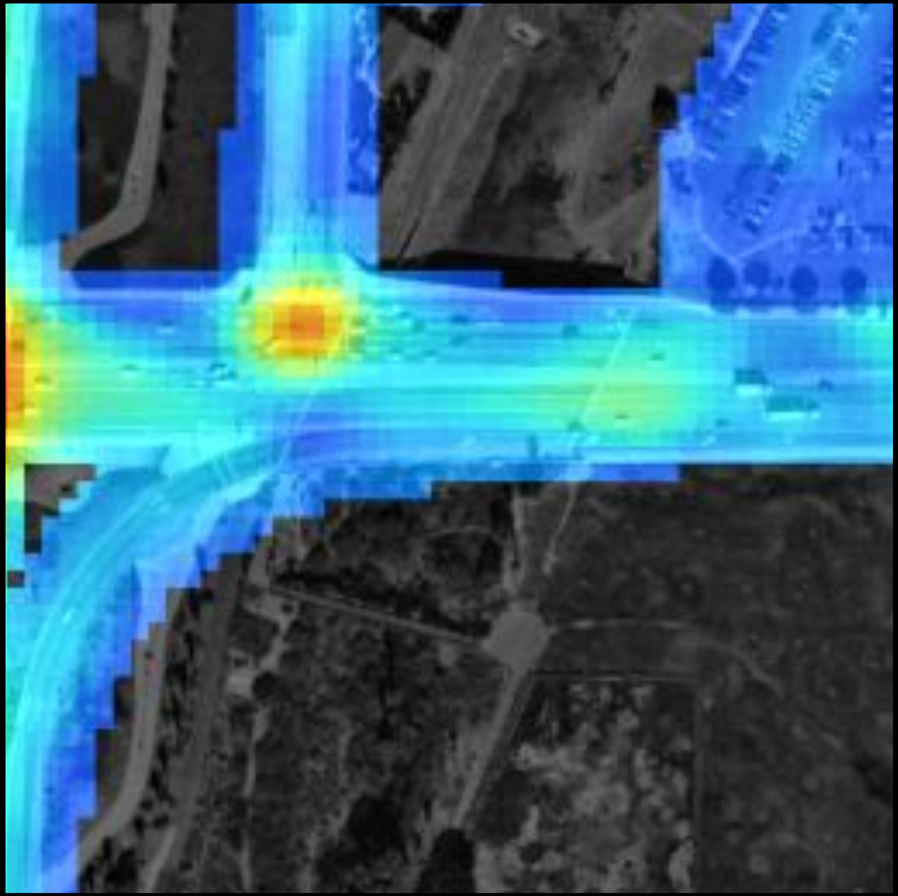
250-Meter Low-Resolution Crash Rate Map (Prior Work)

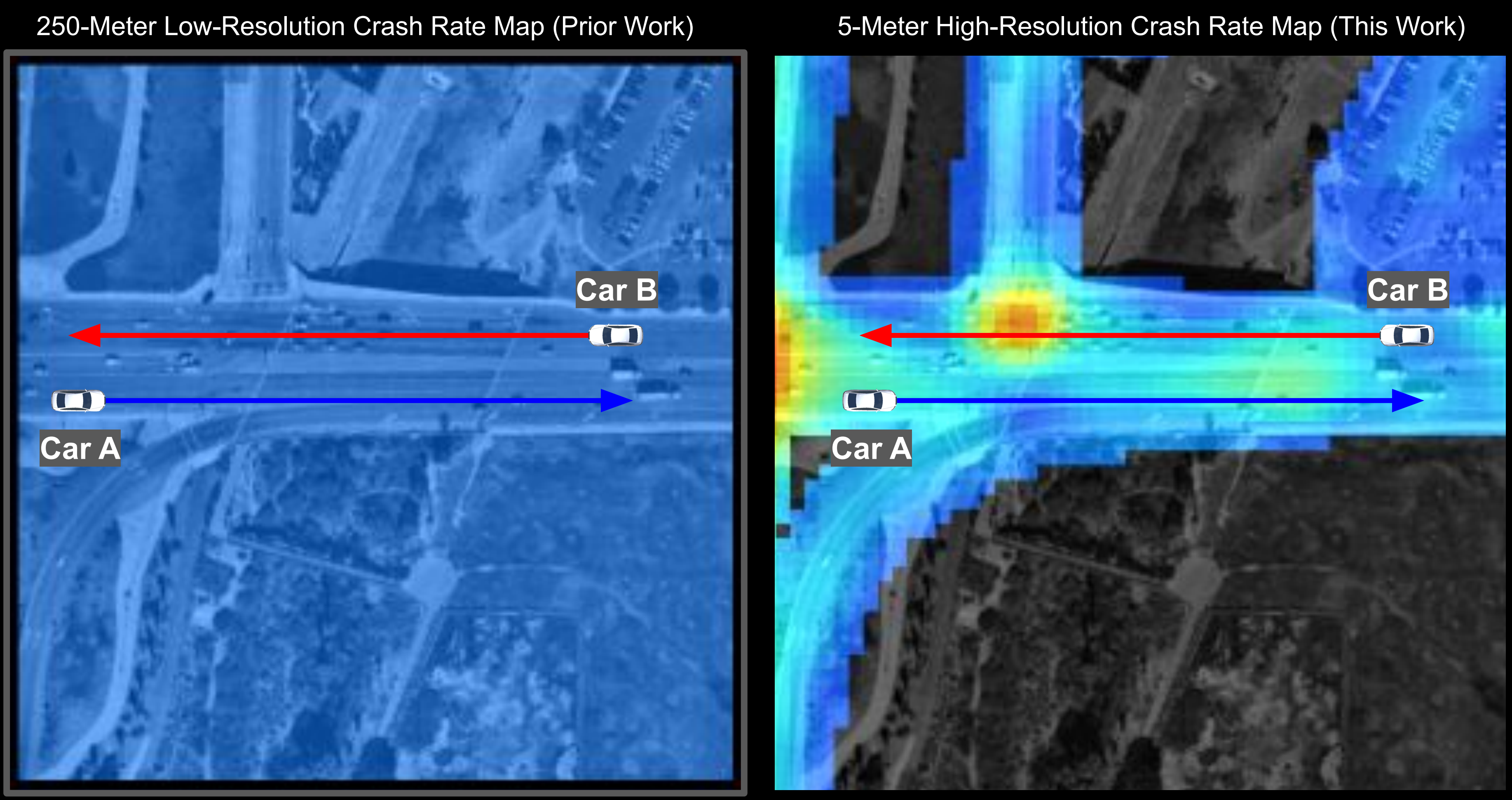
5-Meter High-Resolution Crash Rate Map (This Work)



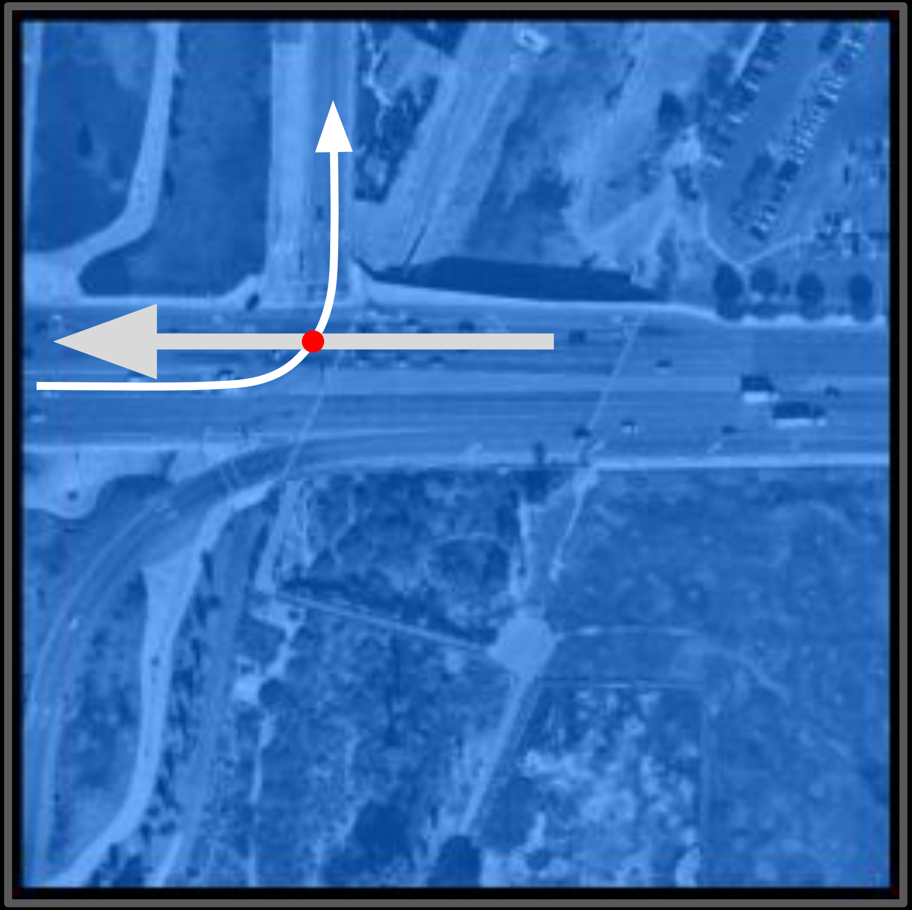
250-Meter Low-Resolution Crash Rate Map (Prior Work)

5-Meter High-Resolution Crash Rate Map (This Work)

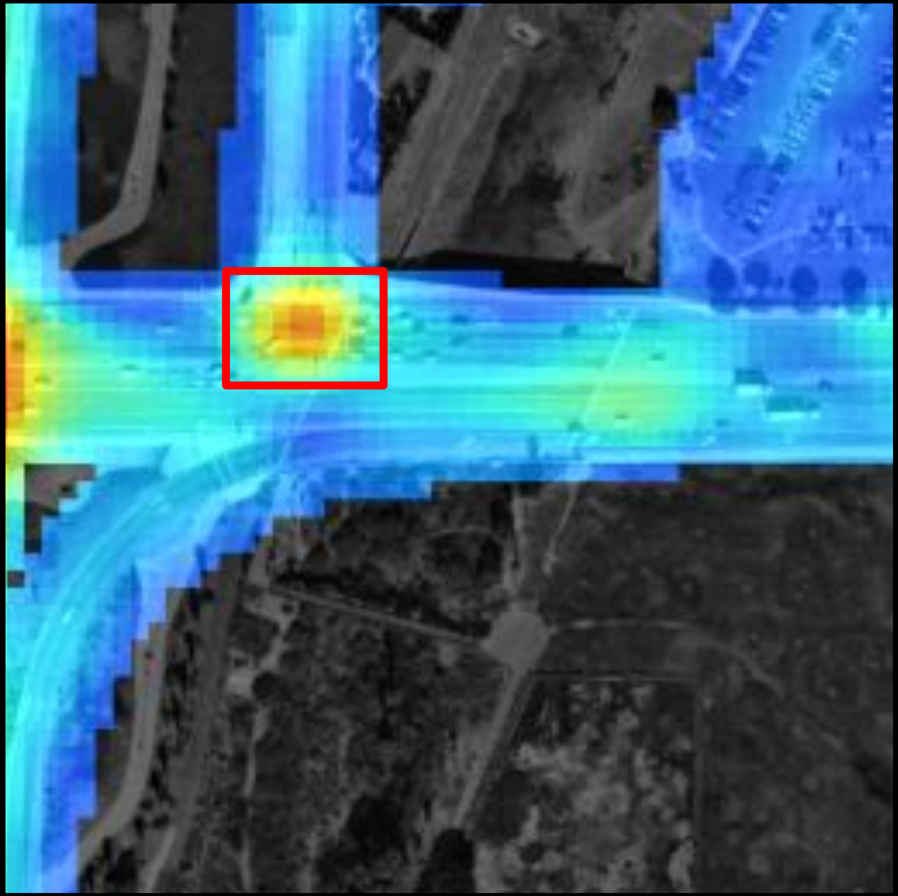



250-Meter Low-Resolution Crash Rate Map (Prior Work)
5-Meter High-Resolution Crash Rate Map (This Work)
Car B
Car A
Car B
Car A

250-Meter Low-Resolution Crash Rate Map (Prior Work)

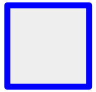


5-Meter High-Resolution Crash Rate Map (This Work)



How Do We Create High-Resolution Crash Rate Maps?

Challenge: Crash data is very sparse



annual rate = 0.1

Basic Solutions:

- Aggregation over time (not practical)
- Aggregation over space (low res.)

Our solution:

- **Aggregation over similar places!**

How Do We Create High-Resolution Crash Rate Maps?

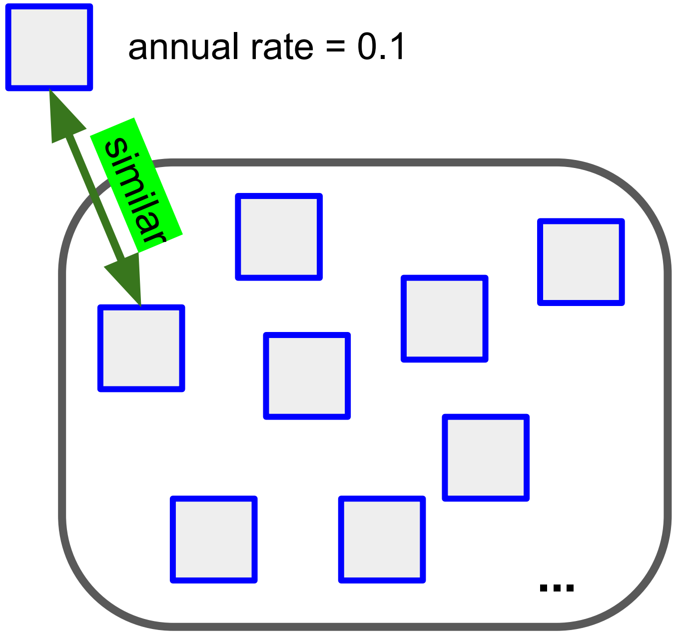
Challenge: Crash data is very sparse

Basic Solutions:

- Aggregation over time (not practical)
- Aggregation over space (low res.)

Our solution:

- **Aggregation over similar places!**



Aggregate observations in 100 similar places

How Do We Create High-Resolution Crash Rate Maps?

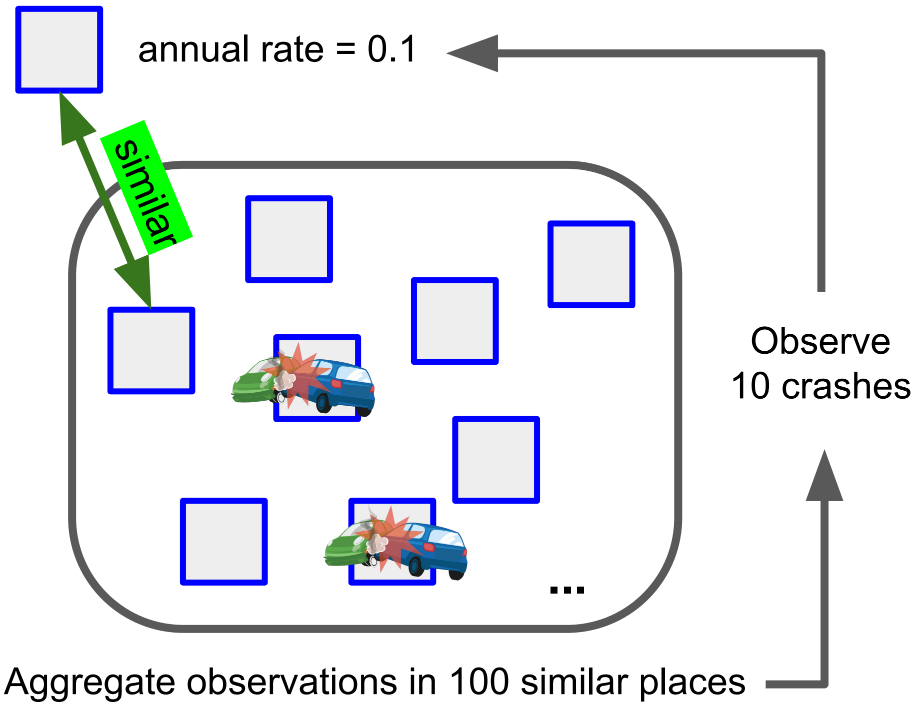
Challenge: Crash data is very sparse

Basic Solutions:

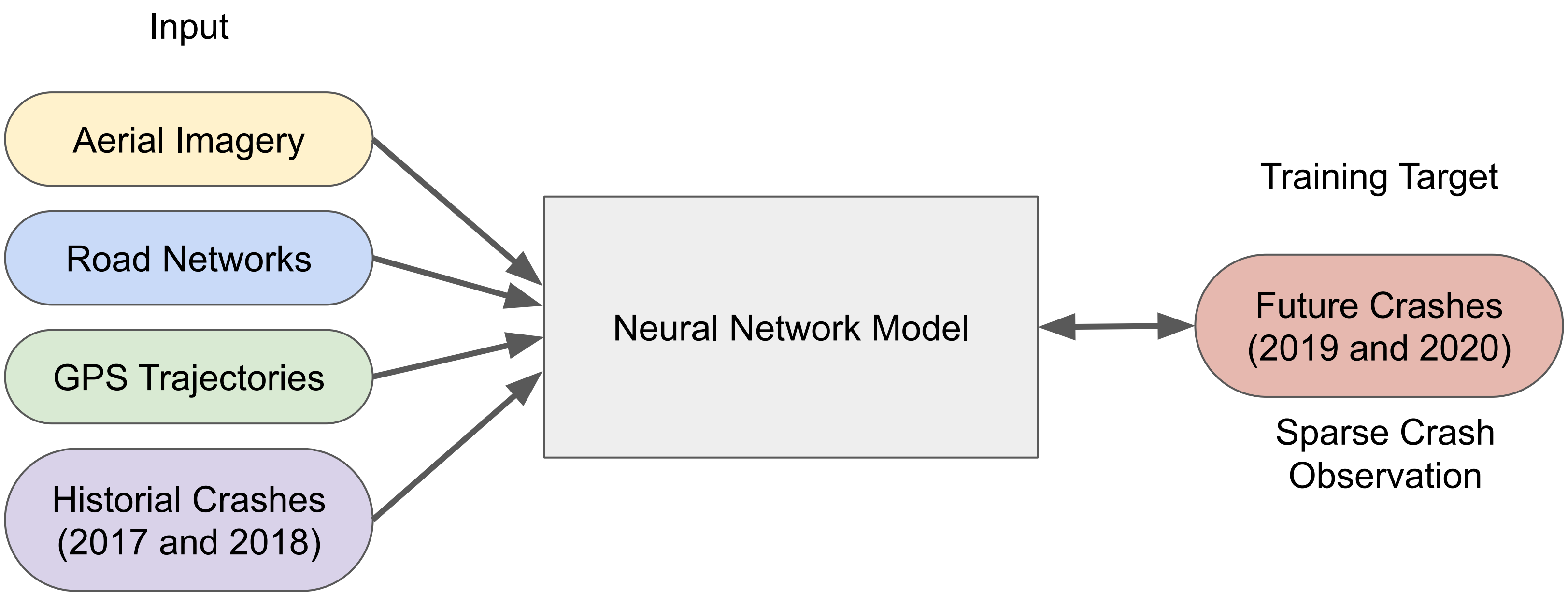
- Aggregation over time (not practical)
- Aggregation over space (low res.)

Our solution:

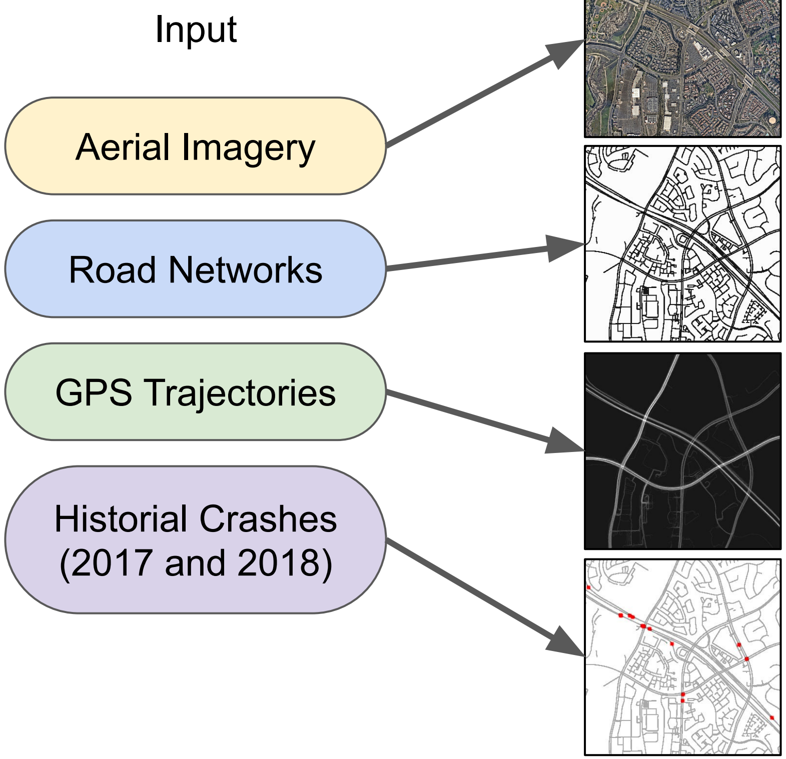
- **Aggregation over similar places!**



Our Solution

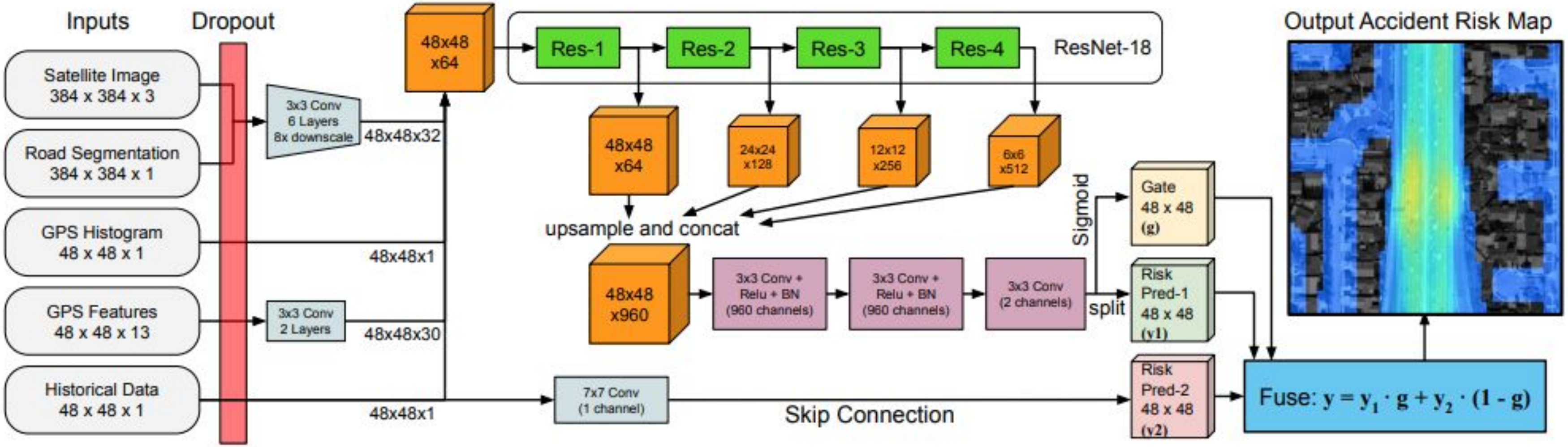


Our Solution - Input



Encode all input data into ‘images’.

Deep Learning Model



Dataset

Total Area: about 7,500 sq km

Total # of Crashes: 517K (four years)

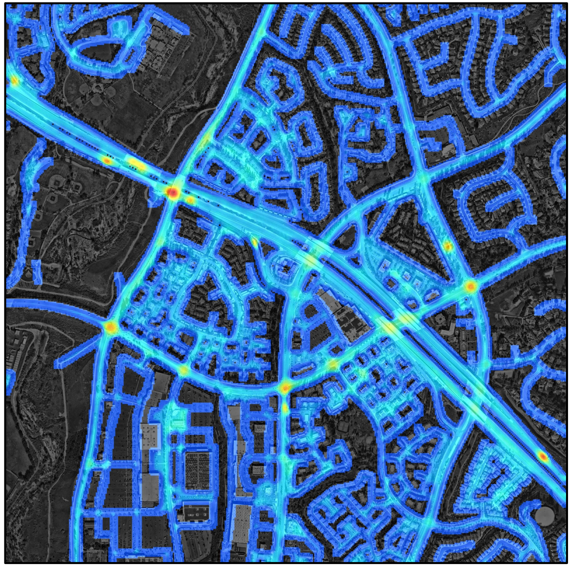
	Los Angeles	New York City	Chicago	Boston
Area (km^2)	3,252	1,832	1,128	1,276
# of Crashes	351K	88K	45K	33K
GPS Traces(km)	3.1M	1.8M	0.7M	2.0M

Accident Dataset Paper:
Accident risk prediction based on heterogeneous sparse data: New dataset and insights (SIGSPATIAL 2019)
Sobhan Moosavi, Mohammad Hossein Samavatian, Srinivasan Parthasarathy, Radu Teodorescu, Rajiv Ramnath

Sources: The US and state departments of transportation, law enforcement agencies, traffic cameras, and traffic sensors within the road-networks.

Evaluation

Crash Rate Map

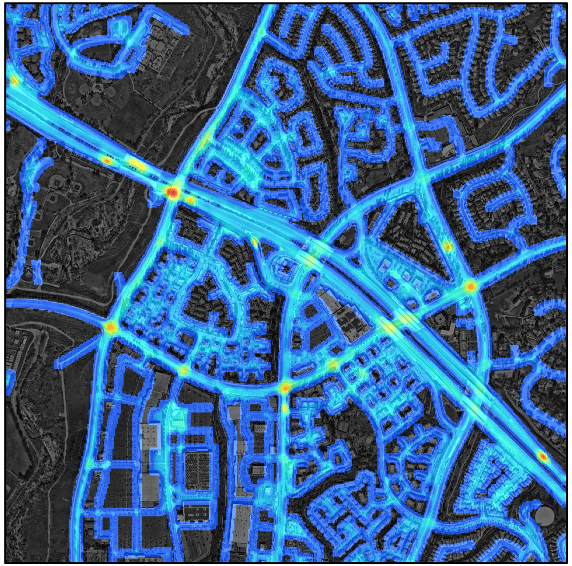


Compare

Future Crashes

Evaluation Metric

Crash Rate Map

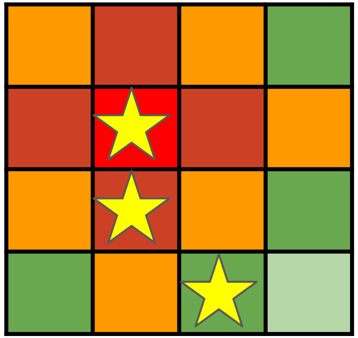


How?

Future Crashes

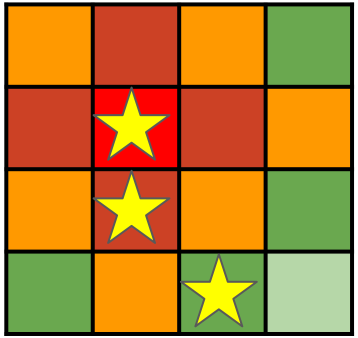
Average Precision Metric

A 4 by 4 Crash Rate Map

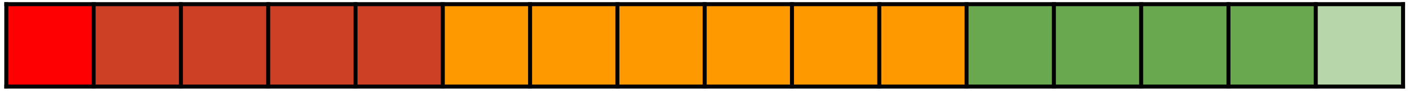


Metric Explanation

A 4 by 4 Crash Rate Map

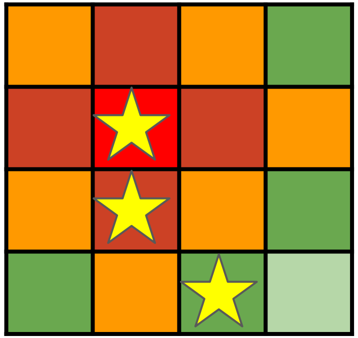


Rank these 16 tiles from high rate to low rate

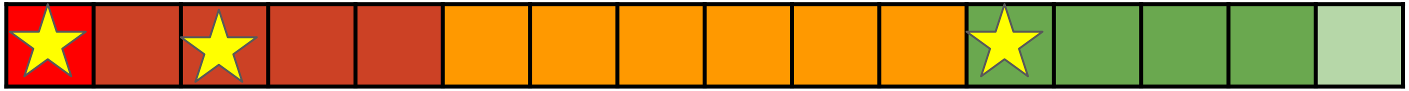


Metric Explanation

A 4 by 4 Crash Rate Map

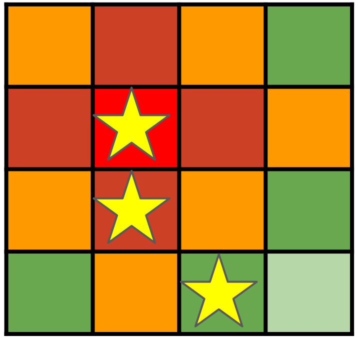


Rank these 16 tiles from high rate to low rate

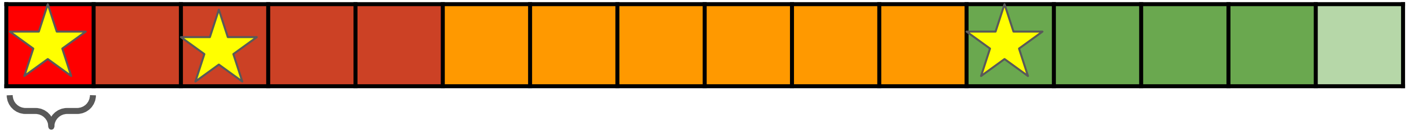


Metric Explanation

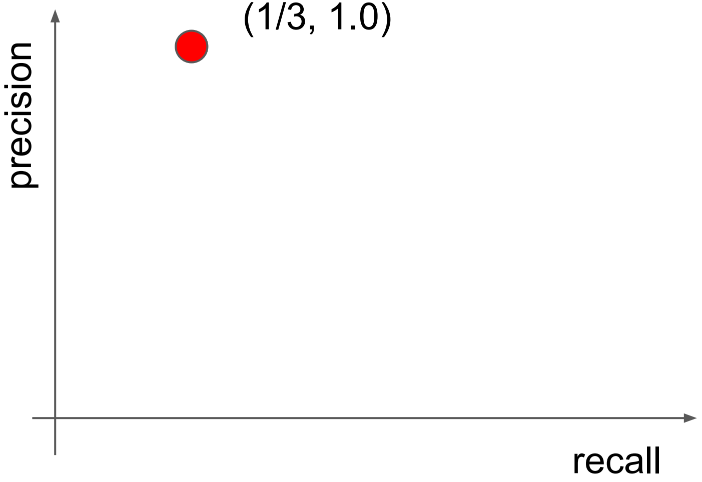
A 4 by 4 Crash Rate Map



Rank these 16 tiles from high rate to low rate

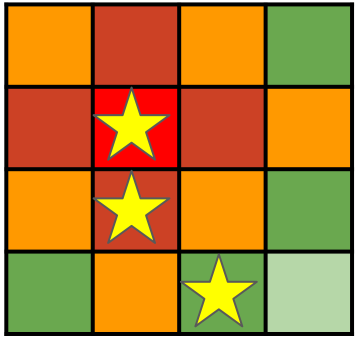


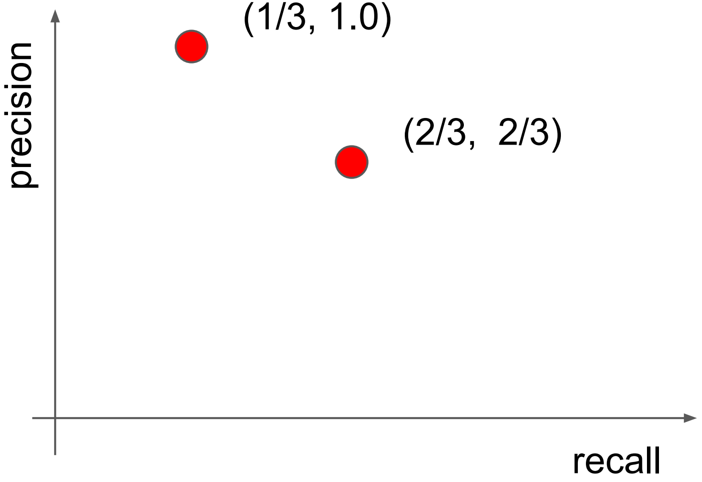
Precision: 1/1
Recall: 1/3



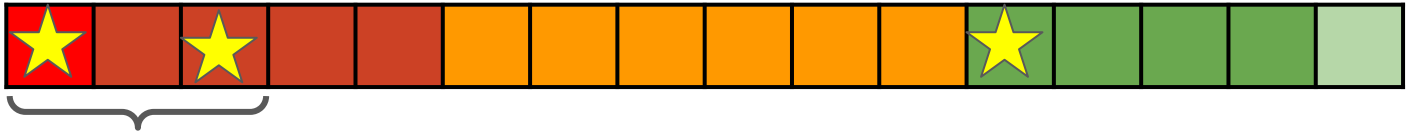
Metric Explanation

A 4 by 4 Crash Rate Map





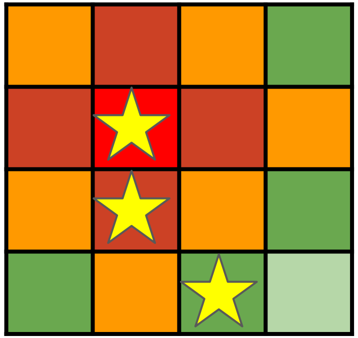
Rank these 16 tiles from high rate to low rate

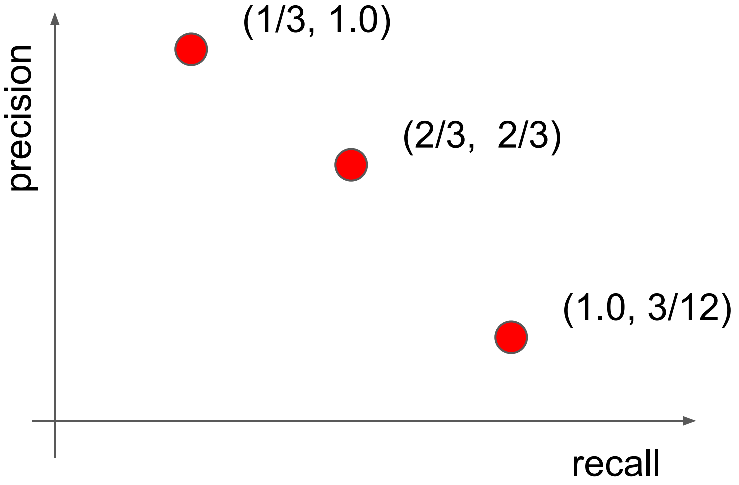


Precision: 2/3
Recall: 2/3

Metric Explanation

A 4 by 4 Crash Rate Map



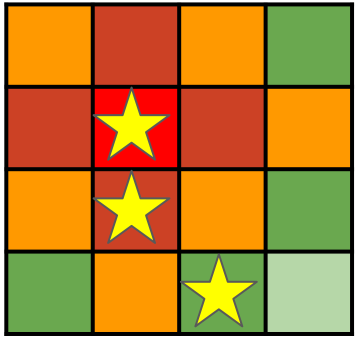


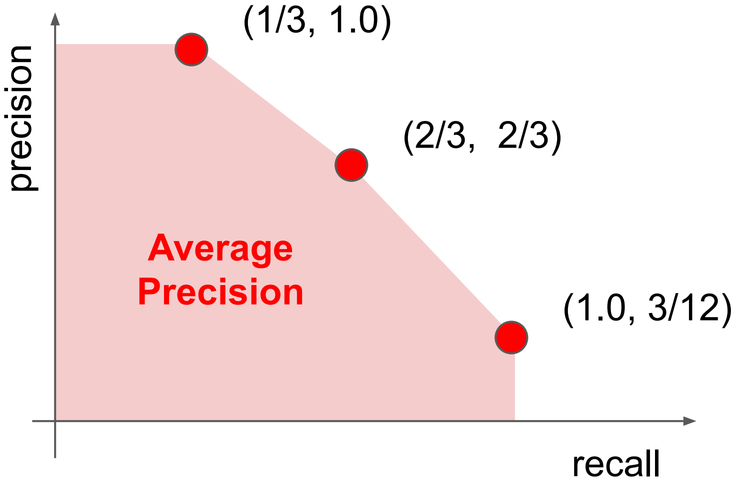
Rank these 16 tiles from high rate to low rate

Precision: 3/12
Recall: 3/3

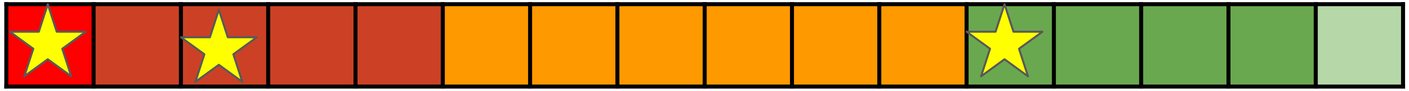
Metric Explanation

A 4 by 4 Crash Rate Map



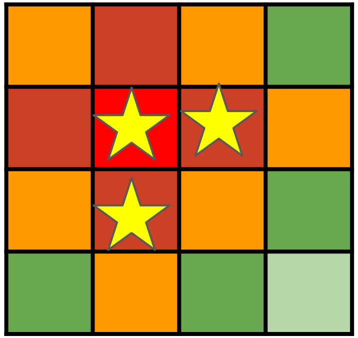


Rank these 16 tiles from high rate to low rate

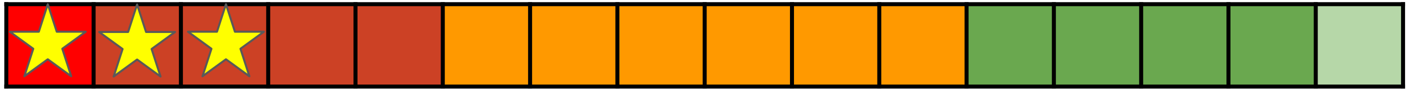


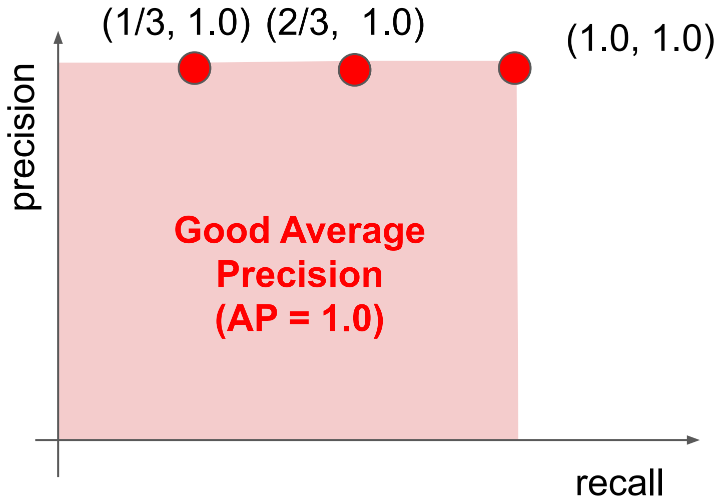
Metric Explanation

A 4 by 4 Crash Rate Map



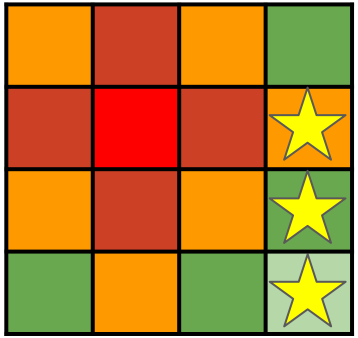
Rank these 16 tiles from high rate to low rate

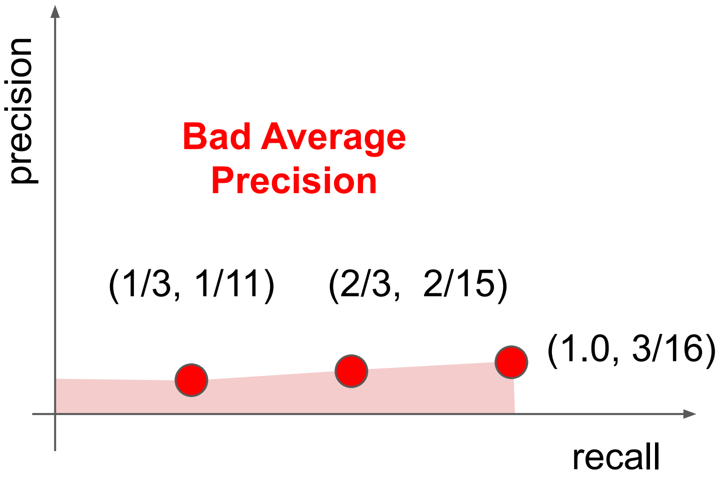




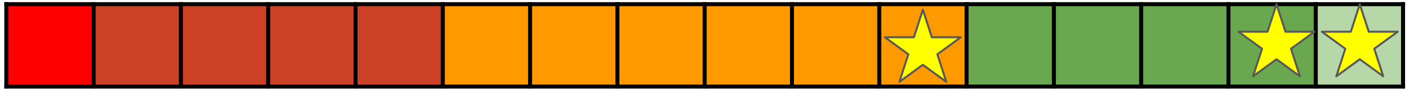
Metric Explanation

A 4 by 4 Crash Rate Map

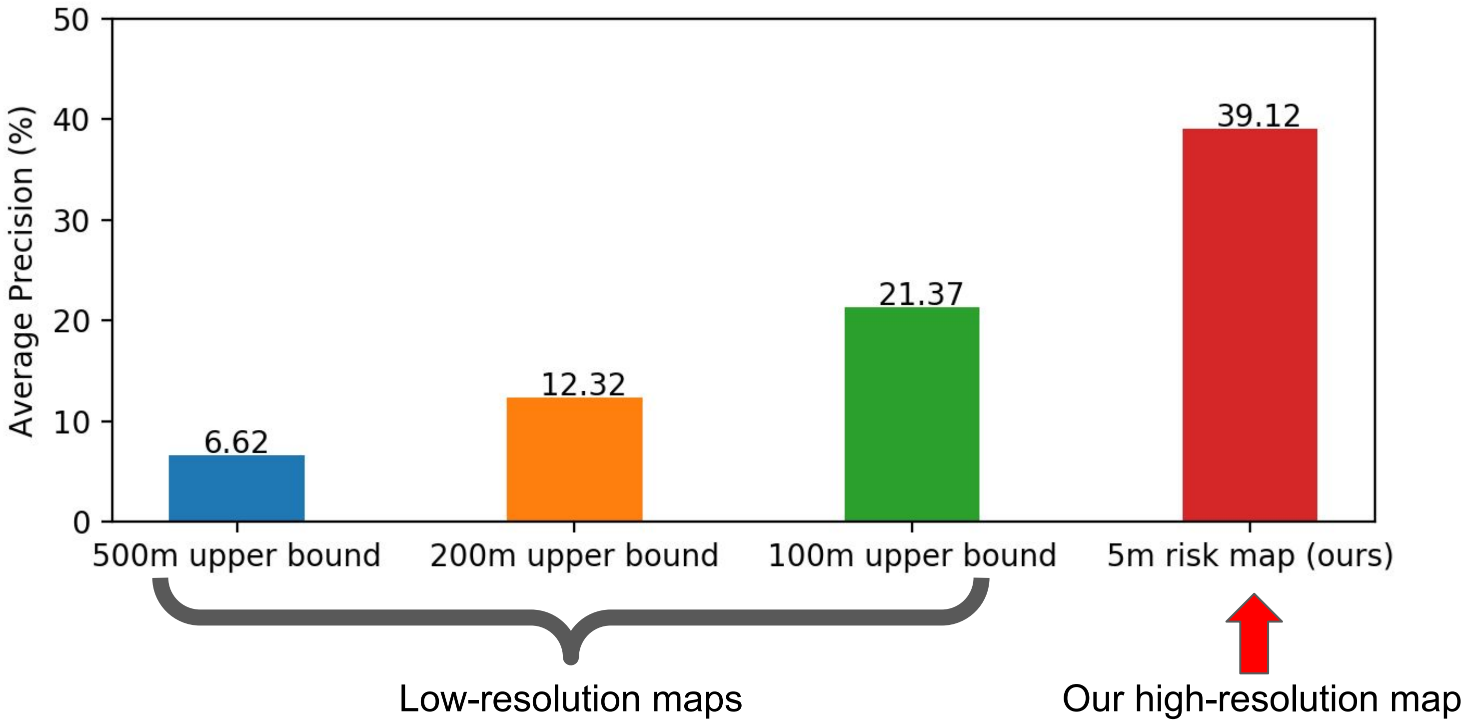




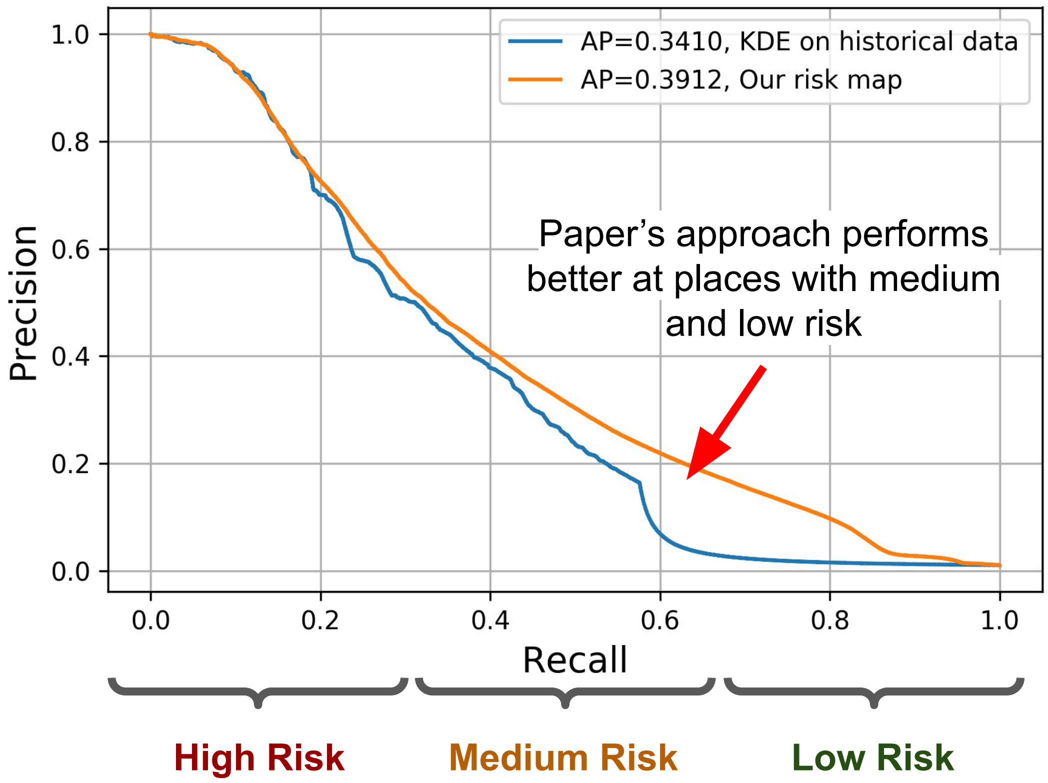
Rank these 16 tiles from high rate to low rate



Performance Versus Low-Resolution Maps



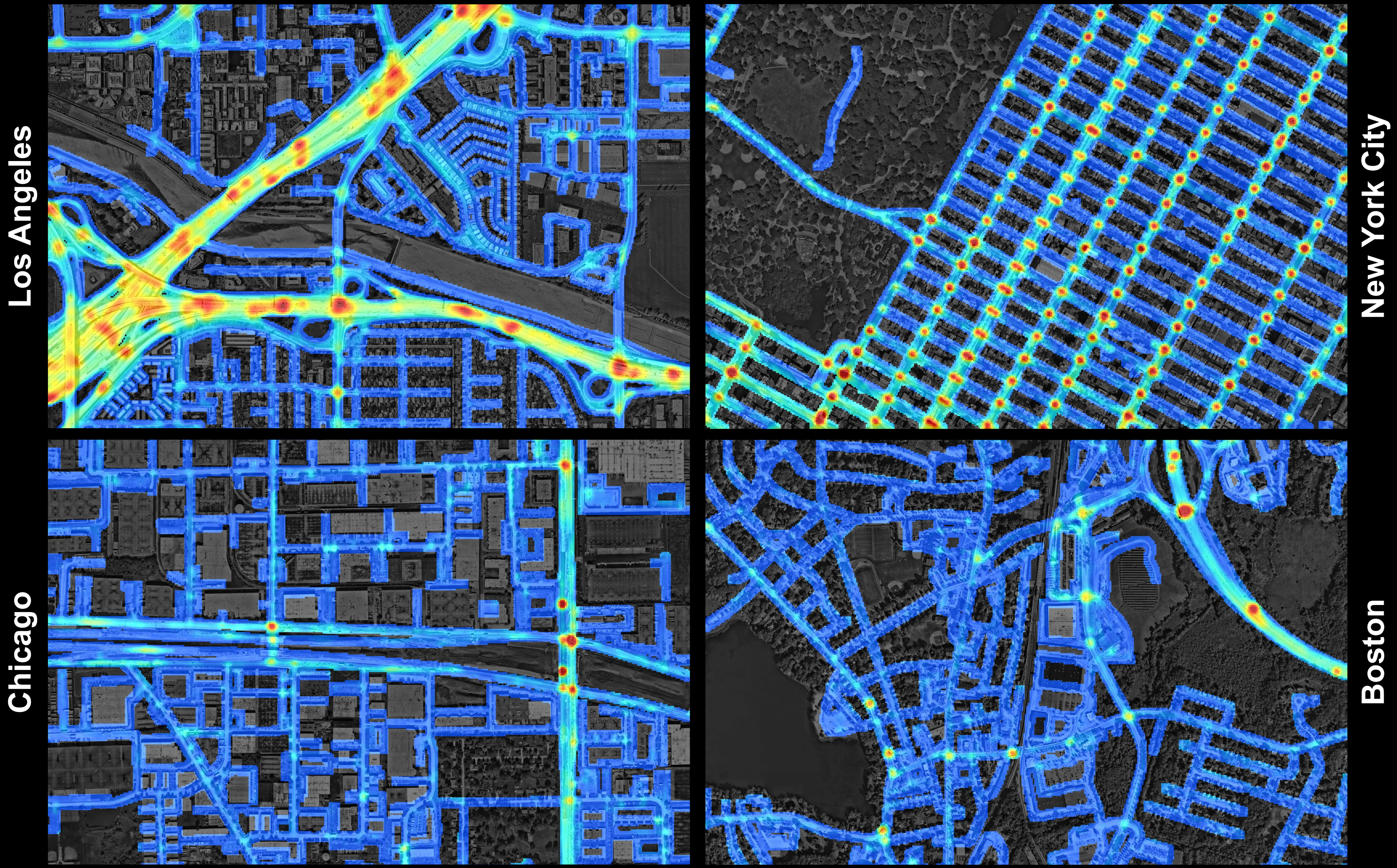
Performance Versus Methods Based on Historical Data



Impact of Different Data Sources

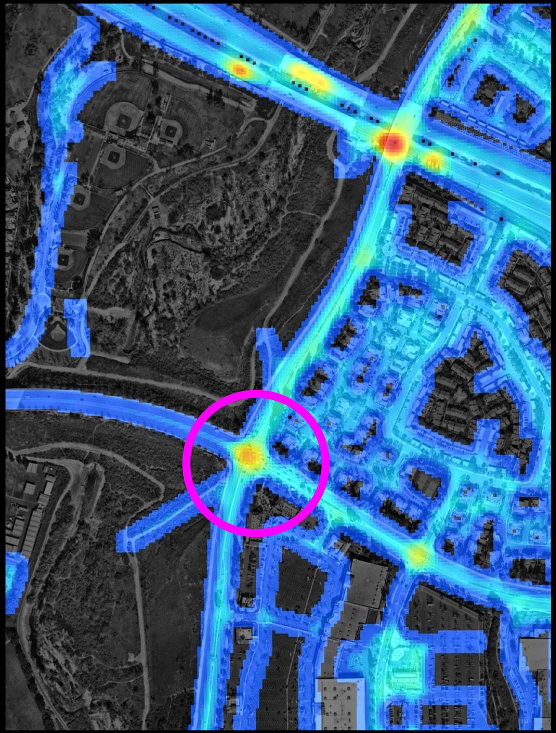
Approaches	Average Precision (%)
Historical Data only (KDE)	34.24
Our approach (road)	35.86
Our approach (road + satellite)	35.10
Our approach (road + GPS)	38.28
Our approach (road + GPS + satellite)	**39.11**

Conclusion: GPS trajectory data is more valuable than the satellite imagery.


Los Angeles
New York City
Chicago
Boston

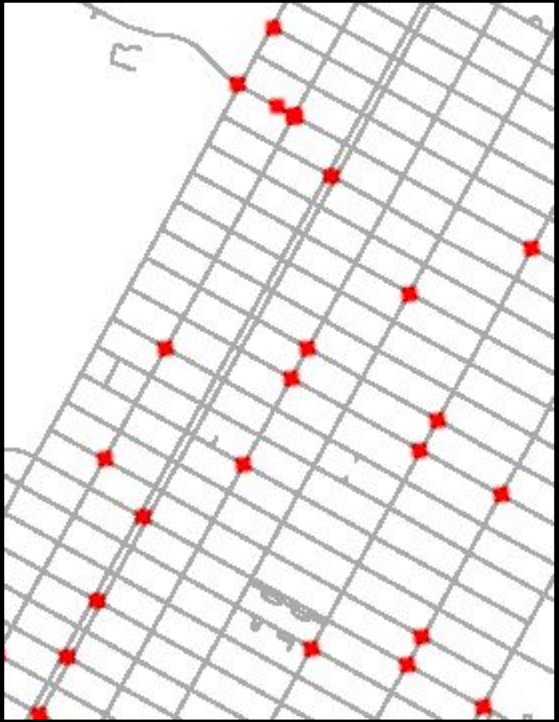
Historical Crashes (2017 and 2018)

Historical Crashes (2017 and 2018)

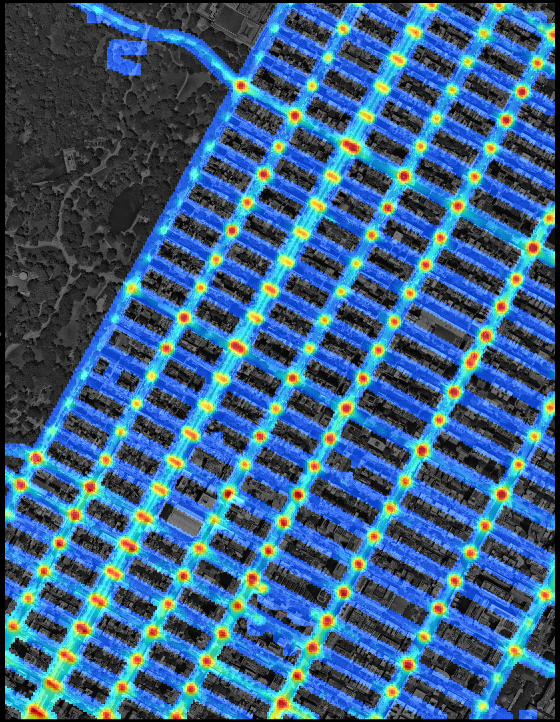


Crash Rate Map

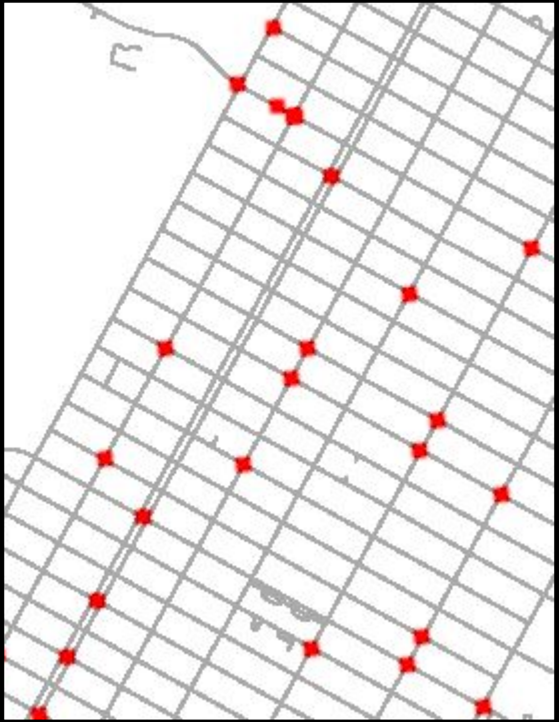

Historical Crashes
(2017 and 2018)
Crash Rate Map
Future Crashes
(2019 and 2020)



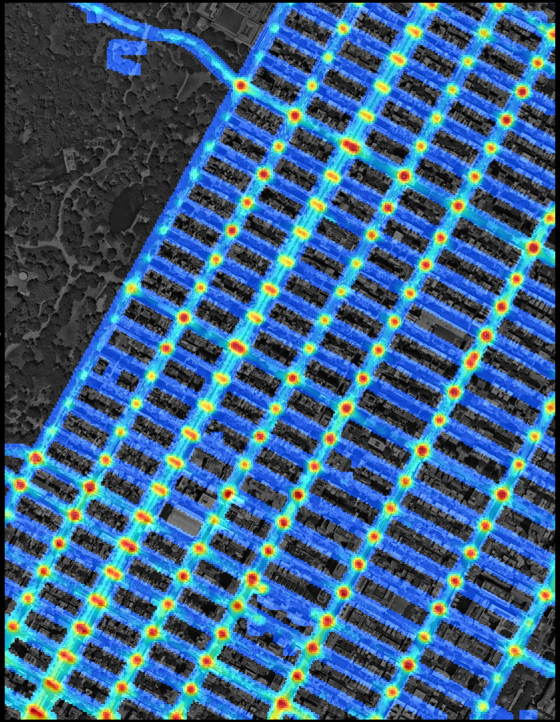
Historical Crashes
(2017 and 2018)



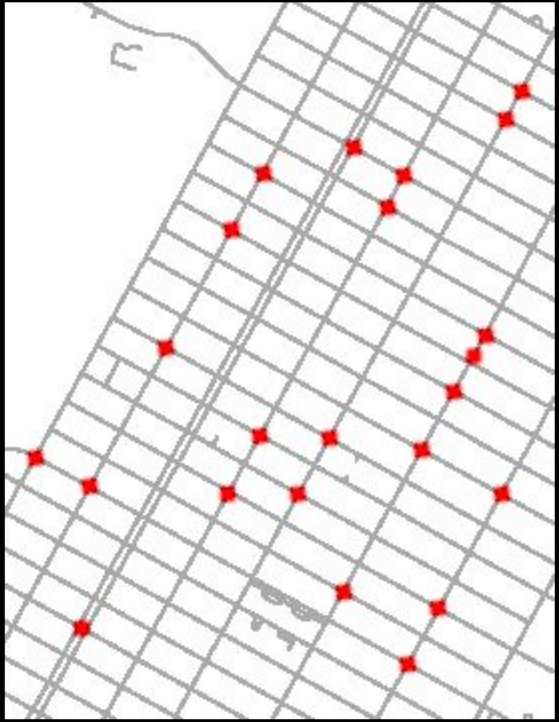
Crash Rate Map



Historical Crashes (2017 and 2018)



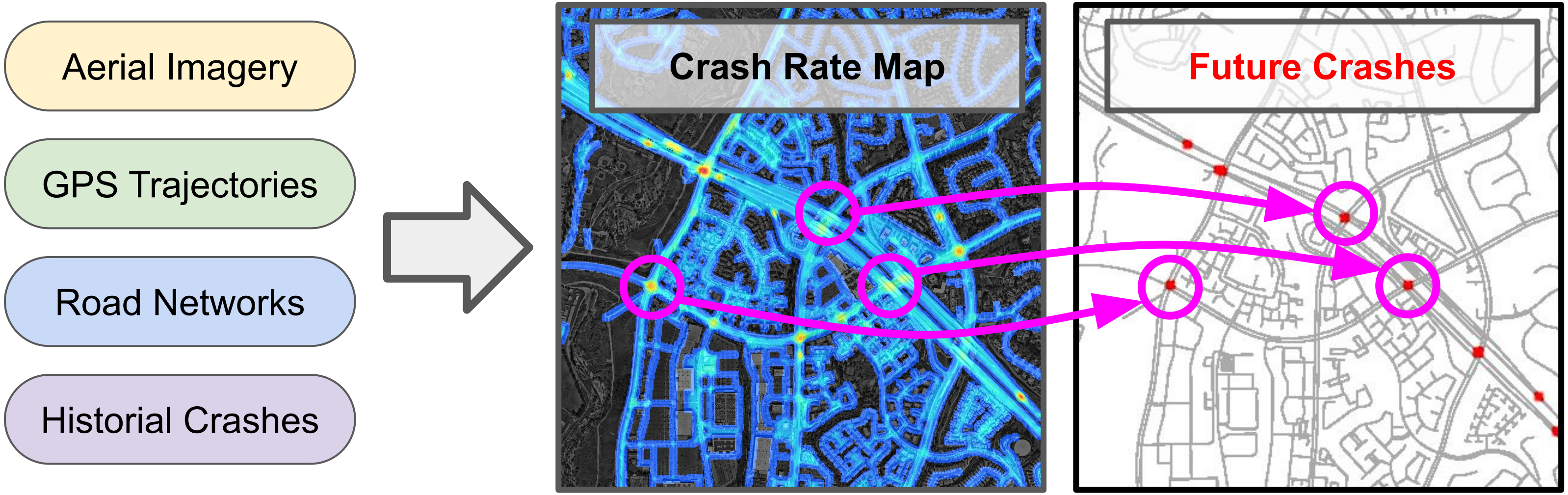
Crash Rate Map



Future Crashes (2019 and 2020)

Traffic Crash Rate Map

- **Contribution**: We demonstrate the possibility of creating high-resolution crash rate maps from sparse crash data.
- **Intuition**: Aggregation over similar places.



Inferring high-resolution traffic accident risk maps based on satellite imagery and GPS trajectories (ICCV 2021)
Songtao He, Mohammad Amin Sadeghi, Sanjay Chawla, Mohammad Alizadeh, Hari Balakrishnan, Samuel Madden