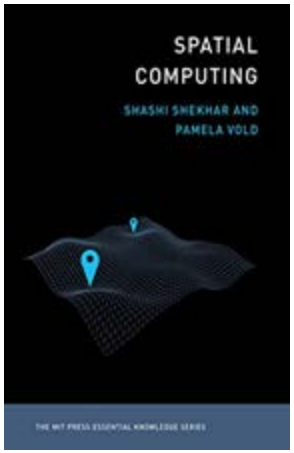
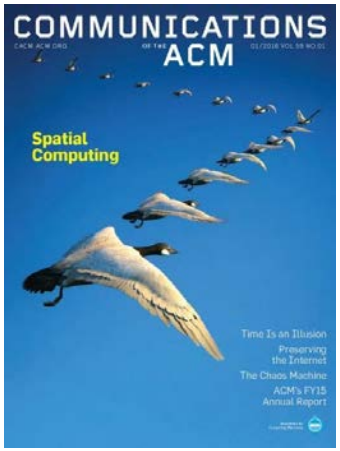
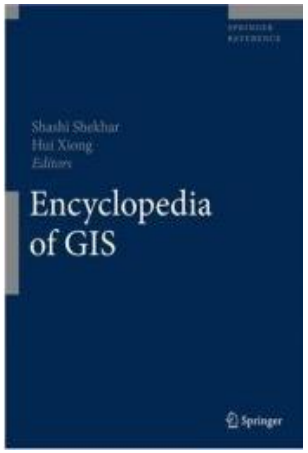
AI and Cities: An International Forum for Innovation and Collaboration (ORANGE)
College of Design, Construction and Planning, University of Florida, May. 9th- 10th, 2025

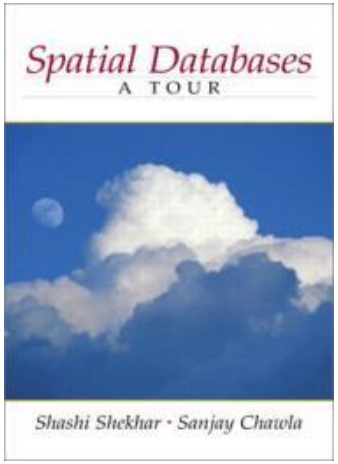
What's Special About GeoAI and Spatial Data Science?

Shashi Shekhar
ADC Chair and McKnight Distinguished University Professor
University of Minnesota
www.cs.umn.edu/~shekhar









Acks.: NSF, USDOD-NGA, USDOE-ARPA-E, USDA-NIFA

Acknowledgements

NSF

- 2118285: HDR Institute: HARP- Harnessing Data and Model Revolution in the Polar Regions
- 2040459: EAGER: Spatiotemporal Big Data Analysis to Understand COVID-19 Effects
- 1901099 : Spatio-temporal Informatics for Transportation Science (8/19 to 7/24)
- 1916518: Midwest Big Data Hub: Building Communities to Harness the Data Revolution
- 1737633: S&CC-IRG Track 1: Connecting **Smart-City** Paradigm with a Sustainable Urban Infrastructure Systems Framework to Advance Equity in Communities

USDA

- 2023-67021-39829: AI Institute for Land, Economy, Agriculture and Forestry (AI-LEAF), $20M, 6/2023 5/2028. AFRI Competitive Project No. MINW-2023-03616
- 2021-51181-35861: Winterturf: A holistic approach to understanding the mechanisms and mitigating the effects of winter stress on turfgrasses in northern climates

USDOD

- IARPA: ROMULUS and REMUS, Two Systems for Building **Mega-City** Intelligence, $900K, HAYSTAC (Hidden Activity Signal and Trajectory Anomaly Detection) Program, 3/2023 3/2027.
- HM0476-20-1-0009: Identifying Aberration Patterns in Multi-attribute Trajectory Data with Gaps

USDOE

- CX-020456: Improving the Freight Productivity of a Heavy-Duty, Battery Electric Truck, USDOE Office of Energy Efficiency and Renewable Energy (FOA DE-FOA-0002044 via Volvo Technology of America LLC), $1.5 M 1/1/2020 - 12/31/2024, (with W. Northop)

NSF 1737633: Connecting the Smart-City Paradigm with a Sustainable Urban Infrastructure Systems Framework to advance Equity in Communities

Major Innovations

- Comprehensive intra-urban scale data (SEIU-EHW parameters, Fig. 1)
- Spatial Data Science to understand relationships
- Model & visualize multi-infrastructure spatial smart city futures
- Knowledge co-production theories, science and practice

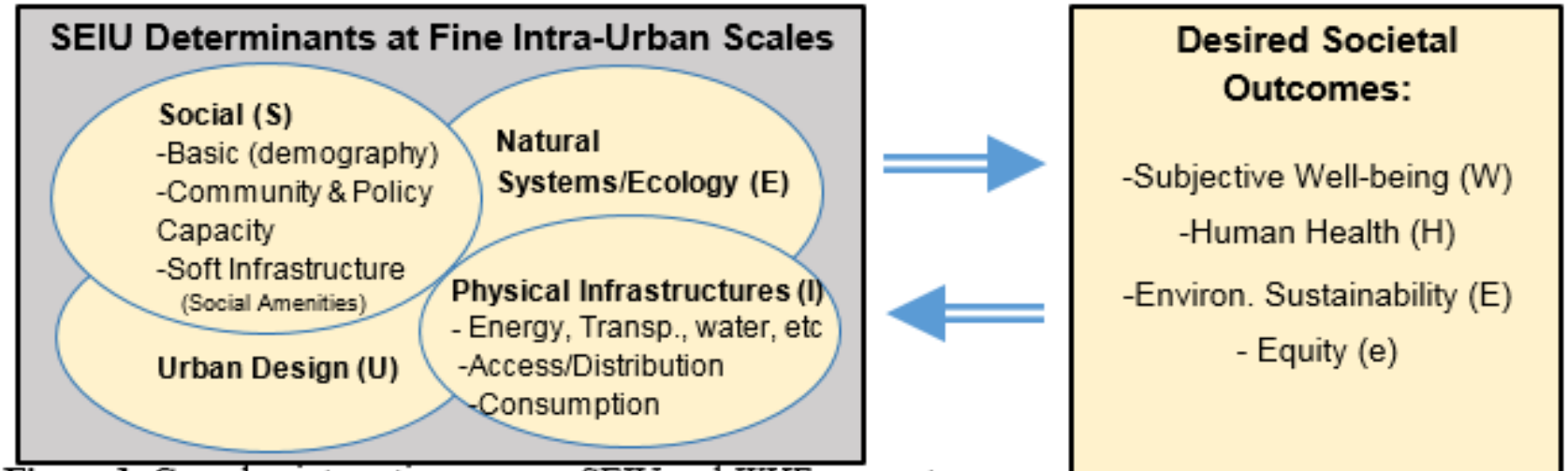


Figure 1: Complex interactions among SEIU and *WHE* parameters.

Community Engagement

- Minneapolis, St. Paul & Tallahassee; Hennepin county, Citizen science with middle and high-schoolers
- MetroLab Network, Natl. League of Cities, ICLEI-USA, Intl. City/County Mgmt. Association

Project Aims

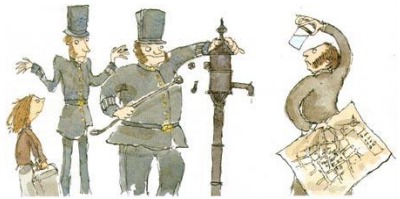
- **Objectives**
  - Understand spatial equity across infrastructures and outcomes (Wellbeing, Health, Environment)
  - Advance WHEe outcomes using smart spatial infrastructure planning in cities.
- **Challenges:**
  1. Data Gaps: need intra-urban scale data on SEIU and WHE parameters (Theme 1)
  2. Knowledge Gaps (Themes 2, 3):
     Data science to understand spatial interactions among SEIU-*WHEe parameters*.
     All-infrastructure models of spatial futures in changing climate, with disruptive infrastructures & technologies
- **Approach in collaboration of Community Partners**

Theme 1: Develop comprehensive data sets on SEIS-EHW at intra-urban scales:
• **Cyber infrastructure** for diverse and disparate data sets
• **Novel citizen science, sensor and survey techniques** to characterize
  ○ air pollution
  ○ near-realtime flooding
  ○ subjective well-being (W)
	Theme 2: Advance spatial data analysis to understand SEIU-*WHEe* relationships
	• Advanced **spatial computing algorithms**
	• Data and Discipline-inspired Hypotheses
	• Equity (e) as spatial dispersion & correlation of *WHE*-SEIU
		Theme 3: Model and visualize spatial smart city futures for Equity-First Plan
		• **Multiple & connected spatial infrastructure futures** scenario modeling
		• Scenario **Visualization**
		• **Value of information** and policy-learning

Theme 4: Education and Workforce Development: Citizen science with middle & high-school students; Interdisciplinary Graduate Certificate; Professional education; Visualization for Policy Leadership;		

An Urban Story: **Geo-Intelligence** and **Spatial Data Science**

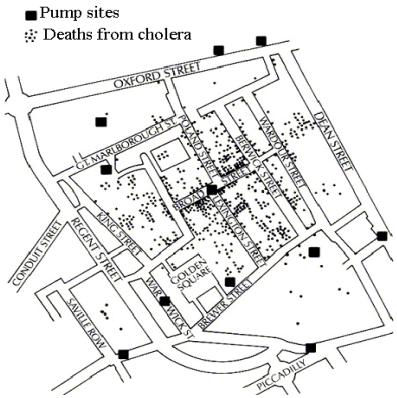
1854: What causes Cholera?



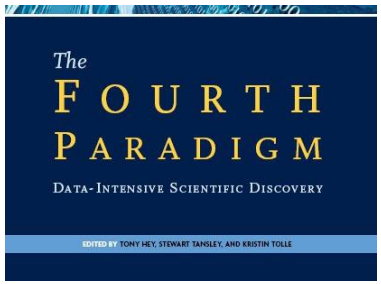
Collect & Curate Data → Discover Patterns, Generate Hypothesis → Test Hypothesis (Experiments) → Develop Theory

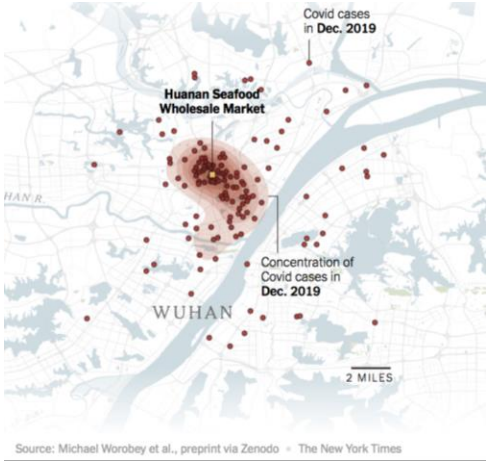
? water pump

Remove pump handle



Impacts: Hygiene, Separate drinking water and sewage systems, …





What has changed?

	Before	Now
Spatial Data Revolution	Smaller data from surveys, few satellites and sensors	**Spatial Big Data** from Nano-satellites, Billions of GPS enabled devices, …
Better and Pl	 	
Spatia Proces		
Spatial Data Science		
Spatial Data Visualization		

Spatial Data Revolution

- Remotely sensed Imagery
  - Thousands of (Nano-)satellites
  - UAVs, Aerial imagery, …
- (GPS-) Location traces
  - Billions of phones, vehicles, …
  - Spatio-temporal patterns of life
- Others
  - Vehicle On-board diagnostics
  - Cellular maps of tissues
  - Geo-social media, Sensor Network, …
- Why is it interesting?
  - See previously inconspicuous
  - Monitor hard to monitor areas
  - Solve previously unsolvable problems

The New York Times

Published: May 13, 2011

McKinsey Global Institute

The study estimates that the use of personal location data could save consumers worldwide more than $600 billion annually by 2020. Computers determine users' whereabouts by tracking their mobile devices, like cellphones.

Urban Garden Detection using Aerial Imagery

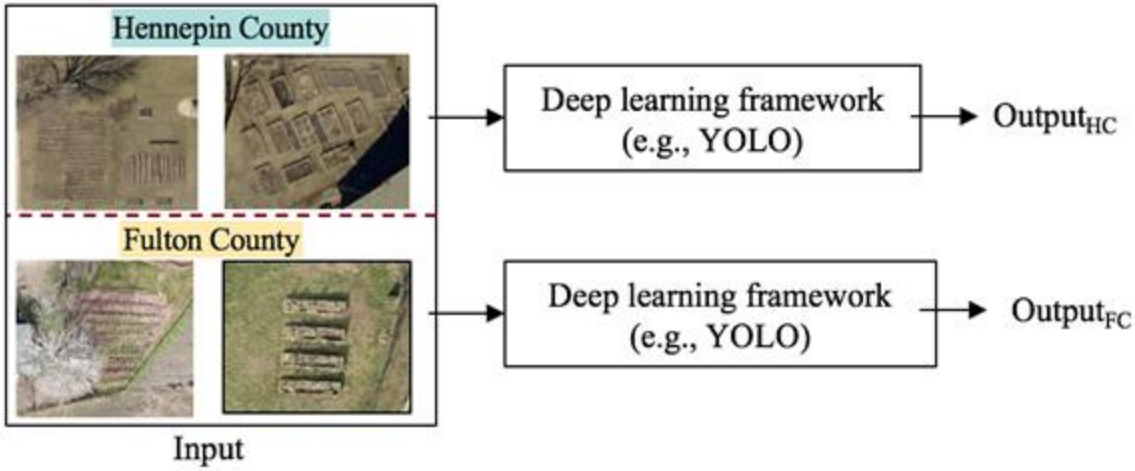
Goal: Assess urban garden food production capacity.

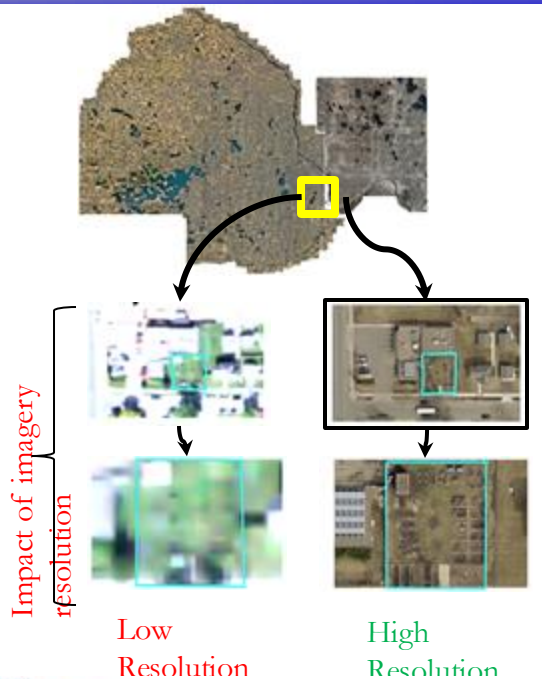
Initial Approach

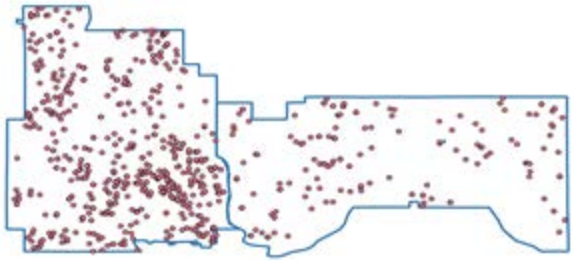
- Digitize urban garden from Google Earth imagery
- Walk neighborhoods for ground-truthing.
- Worked well in Detroit, not in Minneapolis (shadows, trees)

Alternative Approach

- High-resolution (inch-size pixels) spring **aerial imagery**.
- Spatial Variability Aware Neural Networks (SVANN)







	Hennepin County		Fulton County	
	Flat	Raised	Flat	Raised
Train	353	475	12	276
Test	213	273	9	122
Total	566	748	21	398

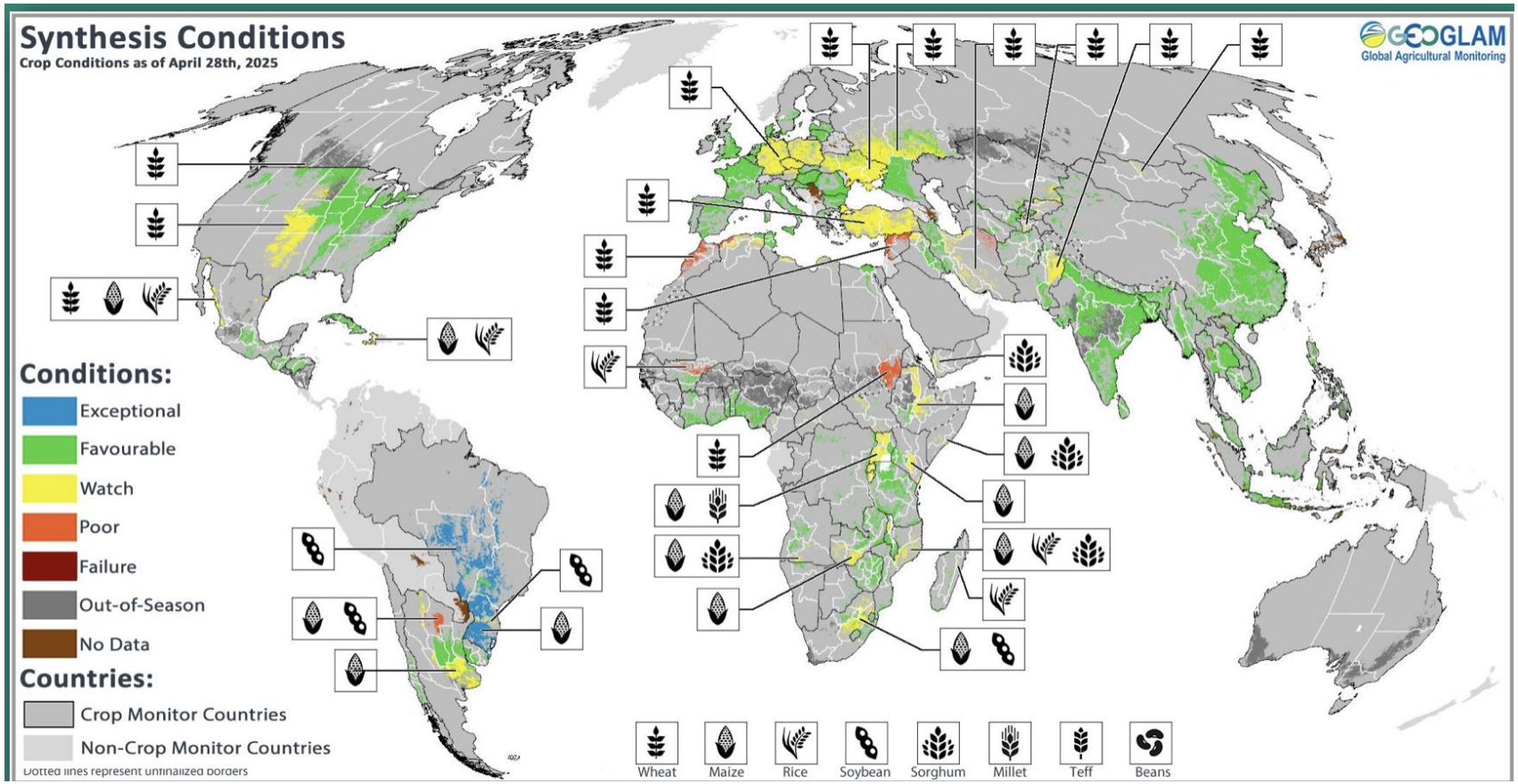
Figure: Train and test dataset

Details:

J. Gupta et al., Spatial Variability Aware Deep Neural Networks (SVANN): A General Approach. ACM Trans. Intell. Syst. Technol. 12, 6, Article 76 (December 2021), 21 pages. https://doi.org/10.1145/3466688

Monitor Global Crops for Early Warning

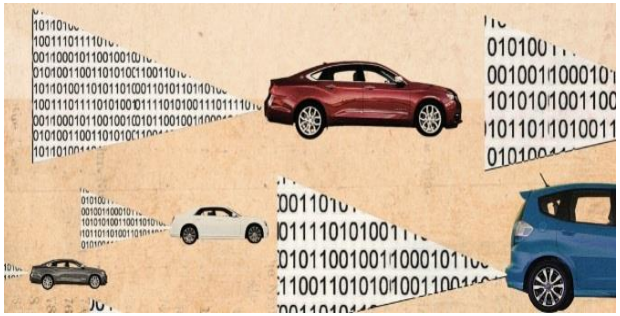
- Last century: US Wetland inventory took 4 decades and $400M
- Now: Monthly global crop-health maps for early warning and action



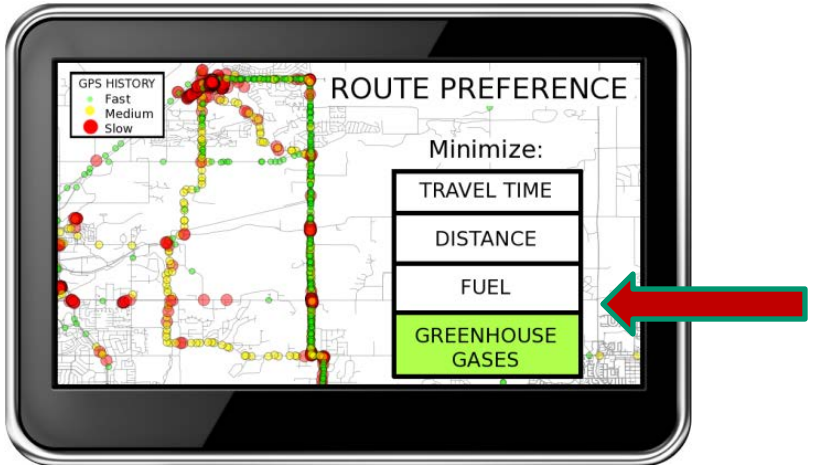
AMIS Agricultural Market Information System

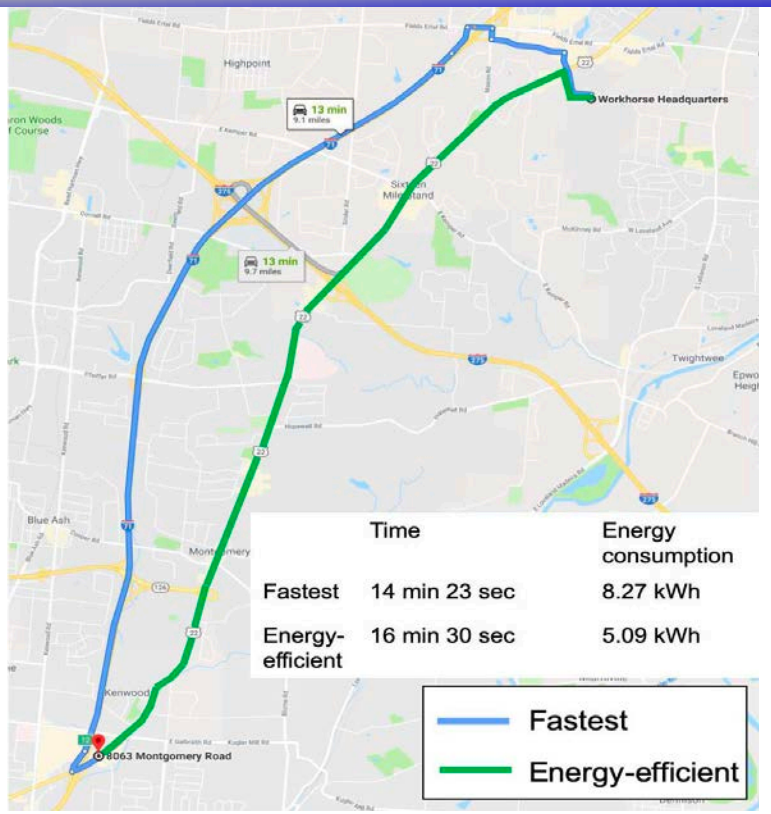
GEOGLAM Global Agricultural Monitoring

Monitor energy use and emissions => Eco-Routing



GPS Tracks + On Board Diagnostics Data





- Oct. 2021: Google Maps supports Eco-routes

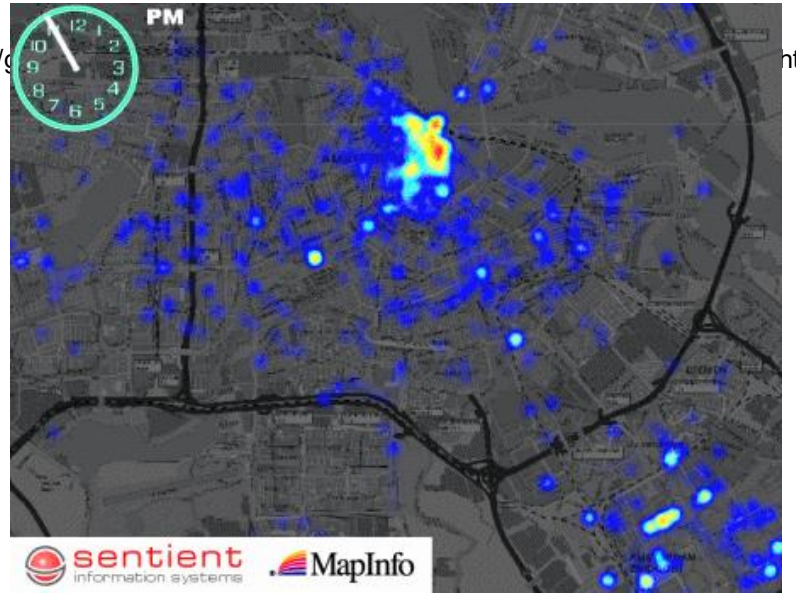
Details: Y. Li, P. Kotwal, P. Wang, Y. Xie, S. Shekhar, and W. Northrop, Physics-guided Energy-efficient Path Selection Using On-board Diagnostics Data, ACM/IMS Trans. Data Sc. 1(3):1-28, Article 22, Oct. 2020. (Initial results appeared in Proc. ACM SIG-Spatial, 2018).

Seeing Spatiotemporal Patterns of Urban Life

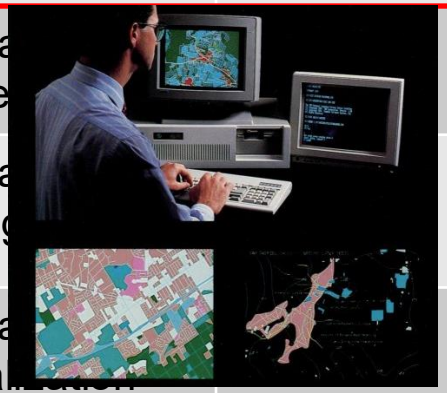
- Google Timelapse
  - Global, 30+ year, 260,000 CPU-hours
  - https://earthengine.google.com/timelapse/
  - Dubai coastal expansion
  - Chicago O'Hare airport
  - Doha, Qatar
  - Marina Center, Singapore (Wikipedia entry)
  - Salt Lake, Bidhannagar, Kolkata, WB, India
  - UMN, Minneapolis;airport, MN, USA
  - Orlando, FL, USA

- Urban crime patterns, Amsterdam

http://g…html



What has changed?

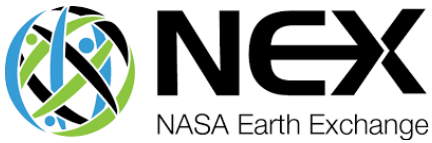
	Before	Now
Spatial Data Revolution	Smaller data	Spatial Big Data
Better Access and Platforms	Smaller platforms, e.g., ESRI Arc/Info, Postgis	Big Compute, e.g., AWS Earth, ESRI GIS Tools for Hadoop,
Spatia… Proce…		
Spatia… Mining		
Spatia… Visual…		

Source: Original PC ARC/INFO Brochure, OCT. 22, 2010 / M. ARTZ, HTTPS://GISANDSCIENCE.COM/2010/10/22/ORIGINAL-PC-ARCINFO-BROCHURE/

Cloud Repositories expand Access to Spatial Big Data

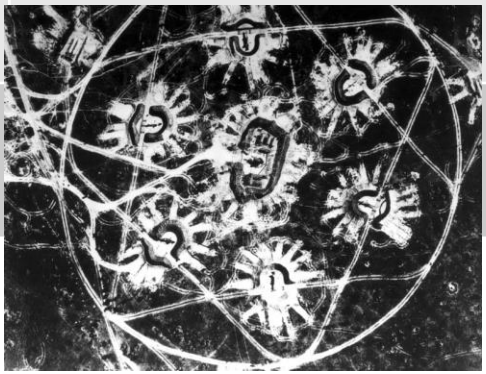
- 2008: USGS gave away 35-year LandSat satellite imagery archive
  - Analog of public availability of GPS signal in late 1980s
- 2017: Cloud-based repositories of geospatial data
  - Explosion in machine learning on satellite imagery to map crops, water, buildings, roads, …

	Google Earth Engines	NEX	AWS Earth
Elevation, Landsat, LOCA, MODIS, NAIP	x	x	x
NOAA	x		x
AVHRR, FIA, GIMMM, GlobCover, NARR, TRIMM, Sentinel-1	x	x	
IARPA, GDELT, MOGREPS, OpenStreetMap, Sentinel-2, SpaceNet (building/road labels for ML)			x
CHIRPS, GeoScience Australia, GSMap, NASS, Oxford Map, PSDI, WHRC, WorldClim, WorldPop, WWF,	x		
BCCA, FLUXNET		x	



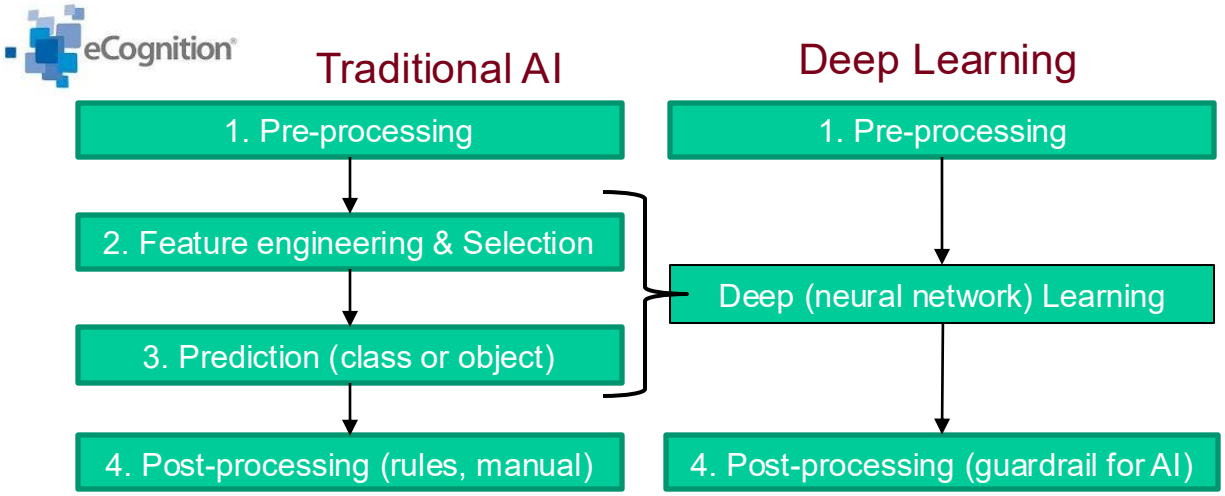
What has changed?

	Before	Now
Spatial Data Revolution	Smaller Data	Spatial Big Data
Spatial Data Access and Platforms	Smaller platforms	Big Compute, e.g., Cloud (e.g., AWS Earth, ESRI GIS Tools for Hadoop)
Spatial Data Processing	Fairly manual, labor-intensive (Geo-Intelligence)	More automation (Geo-Augmented-Intelligence (Geo-AI))
Spatial Data Science		
Spatial Data Visualization		



Traditional AI vs. Deep Learning

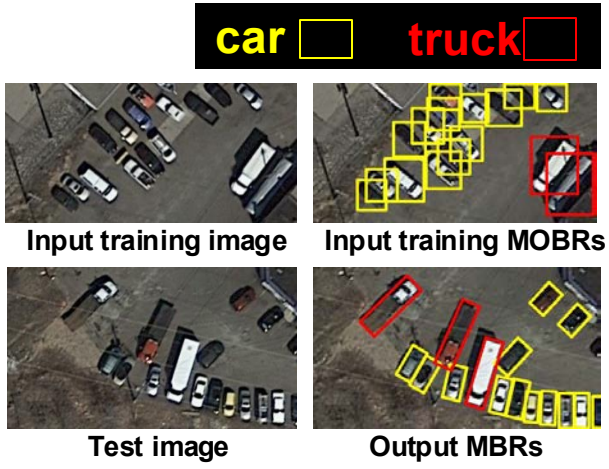
- From Satellite Imagery: Classify Land-cover, Map buildings
- Ex. 2009 Haiti Earthquake: Map building damage [1]



[1] J. Aardt et al., Geospatial Disaster Response during the Haiti Earthquake: A Case Study Spanning Airborne Deployment, Data Collection, Transfer, Processing, and Dissemination, Photo. Eng. & Remote Sens., 77(9):943-952, Sept. 2011.

Deep Learning for Geo-Object Detection

- **Q:?** How many vehicles in a parking lot? City?
- **Ex.:** Estimate truck supply in a city (CH Robinson).
- **Old Computer Vision workflow**
  - Many steps, each adds error
- **New Deep Learning Workflow – fewer steps**
  - Aerial imagery (3 inch pixels, Twincites, MN, USA)
  - NAIP Imagery (1 meter pixels, 2017)
    - MA Buildings Data (https://www.cs.toronto.edu/~vmnih/data/)
- Detected Geo-objects
  - Cars, trucks, Houses, …
  - Method: Convolutional Neural Networks (YOLO)

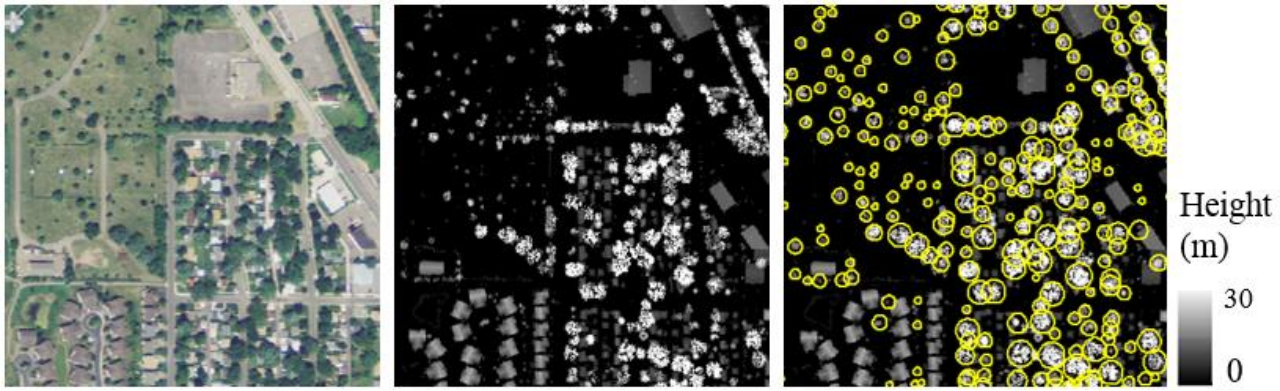


Note: NAIP = National Agriculture Imaging Program (USDA)

Details: Y. Xie et al., An Unsupervised Augmentation Framework for Deep Learning based Geospatial Object Detection: A Summary of Results, Proc. 26th ACM SIGSPATIAL Intl. Conf. Adv. in GIS, 2019.

Mapping Urban Green Infrastructure Trees from Remote Sensing Imagery

- Why?: Protect Powerlines, Manage Emerald Ash Borer, Green infrastructure Equity
- Input: LiDAR + Remotely Sensed **Aerial** Imagery + (NAIP Ground Truth)
- Approach: Tree Inference by Minimizing Bound-and-band Errors (TIMBER)
  - Optimization to find tree locations and sizes
  - Deep learning constructs features separating trees & non-trees (e.g., light pole)



Details: **(a)** Revolutionizing Tree Management via Intelligent Spatial Techniques,
Proc. 27th ACM SIGSPATIAL Intl. Conf. on Adv. in GIS, Nov. 2019 Pages 71–74 (Best Vision Paper).

(b) TIMBER: A Framework for Mining Inventories of Individual Trees in Urban Environments using Remote Sensing Datasets,
IEEE Intl. Conf. on Data Mining (ICDM), 2018..

Generative AI and Urban Planning

- Challenge: Deep Learning is data and compute hungry
  - However, many domains do not have enough labelled data
- An Approach:
  - Train Foundation Models with unlabeled data using self-supervision
  - Q? May knowledge (e.g., mechanisms, laws of geography) also help?

Domain	Foundation Models	Specialized (Downstream) Tasks	Downstream Urban Tasks
Text, NLP	BERT, GPT-x	ChatGPT, Co-Pilot	? Urban Q/A
Weather	Pangu-Weather	Medium range (2hr - 2 weeks)	? Predict urban floods
Geospatial	HLS Geospatial FM	Detect burns, floods, …	? Urban sprawl, …
Urban	trajectory FM	Spatiotemporal patterns of urban life	

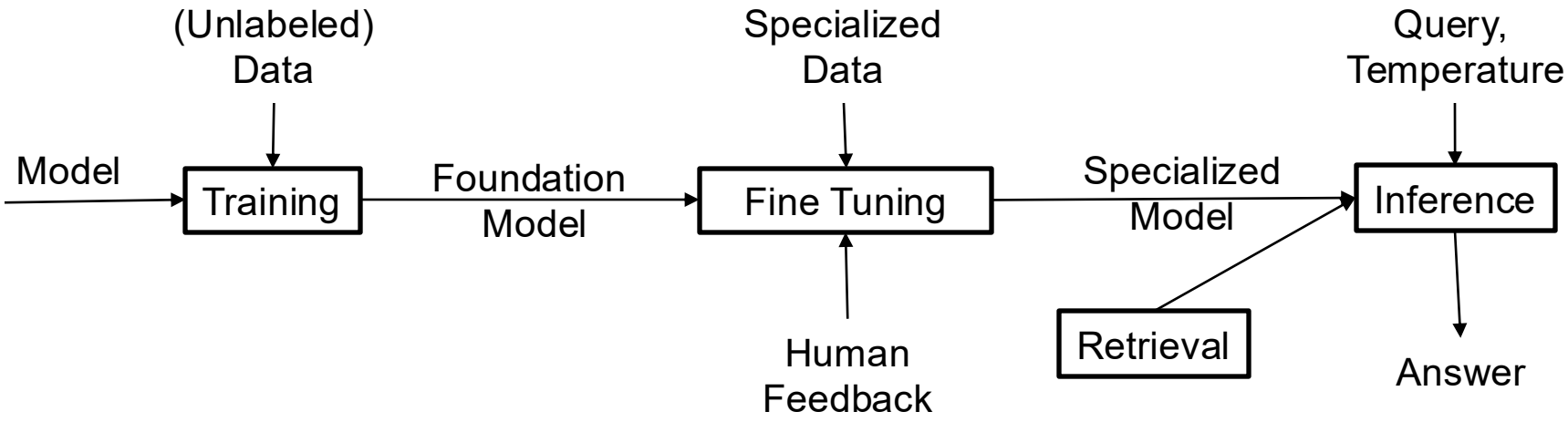
Note: Other options include Diffusion models, Autoregressive Models, Normalizing Flows, GANs

[1] S. Bond-Taylor, A. Leach, Y. Long and C. Willcocks, “Deep Generative Modelling: A Comparative Review of VAEs, GANs, Normalizing Flows, Energy-Based and Autoregressive Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):7327-7347, Nov. 2022. 10.1109/TPAMI.2021.3116668.

[2] Miquel Noguer I. Alonso, The Mathematics of Modern Generative Modeling: Normalizing Flows, Autoregressive and Diffusion Models, Jan. 09, 2025. http://dx.doi.org/10.2139/ssrn.5089573

Foundation Models and Generative AI: Basics

- Background
  - Definition: A foundation model is any model (a) trained on broad data, (b) adaptable to a wide range of downstream tasks [1]
  - Benefits to downstream tasks: Lower sample, computational and energy complexity
  - An architecture [2]



Sources:
[1] R. Bommasani et al, On the Opportunities and Risks of Foundation Models, July 2022, https://arxiv.org/abs/2108.07258
[2] adapted from a presentation by Bill Daily on Government, University, Industry Cooperation: The NVIDIA story at the CCC Computing Futures Symposium, May 15th-16th, 2022.

Generative AI and Spatial Reasoning

Summary: Spatial reasoning and comprehension in (multimodal) LLM is not robust,
But: it is core to commonsense understanding of the world and indeed many domains.
- Tony Cohn, AAAI-2025 Invited Talk on Can Large Language Models Reason about Spatial Information?

I am walking Southeast. I look down on my compass. Please give me a picture of the compass.

Here is an image of a compass viewed from above, with the needle pointing Southeast. Let me know if you need any modifications!

ChatGPT 4o January 2025

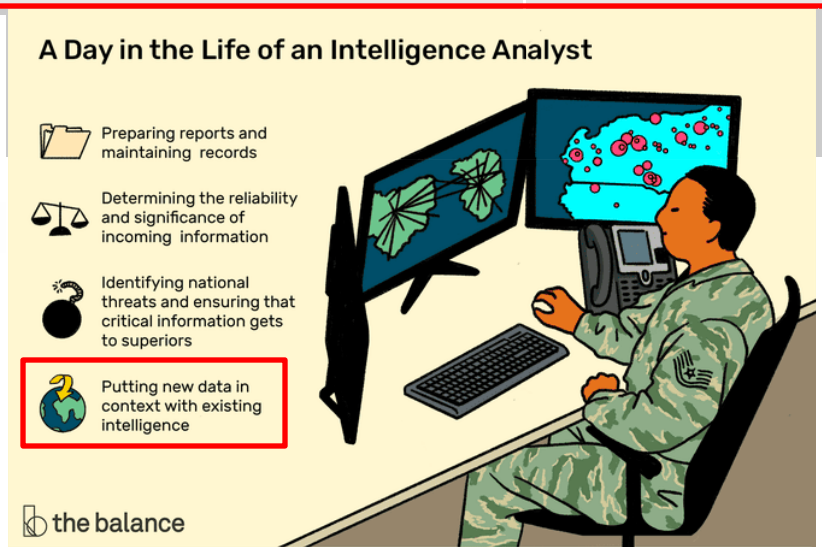
Give me a picture of a voltmeter showing six volts on a twelve volt scale.

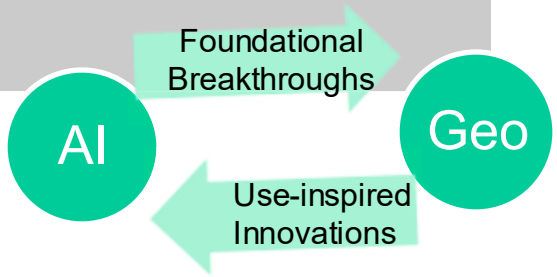
ChatGPT 4o January 2025

Here is an image of a compass viewed from above, with the needle pointing Southeast. Let me know if you need any modification.

What has changed?

	Last Century	This Century
Spatial Data Revolution	Smaller Data	Spatial Big Data
Spatial Data Access and Platforms	Smaller platforms	Big Compute, e.g., Cloud (e.g., AWS Earth, ESRI GIS Tools for Hadoop)
Spatial Data Processing	Fairly manual, labor-intensive (Geo-Intelligence)	More automation Geo-Augmented-Intelligence (Geo-AI)
Spatial Data Science	"One-size fit all" AI applied to spatial data	Virtuous cycle between Geo and AI





Why go beyond Object Detection?

Source: A. Karpathy (2012), The state of Computer Vision and AI: we are really, really far away
http://karpathy.github.io/2012/10/22/state-of-computer-vision/

Need for AI (Vision) to break out of "RGB+Lidar" box

A downpour in Las Vegas at the CES technology show a year and a half ago may prove to have been a watershed moment in the race to develop autonomous cars. While other companies promoting their experimental self-driving vehicles had to keep them parked in the rain, one company, AdaSky, demonstrated how its sensors could see people hundreds of feet ahead even in a downpour, and even when they were wearing white and standing against a white background.

"Thermal imaging is the best sensor at detecting people, day or night," Chris Posch of FLIR Systems said.

These High-Tech Sensors May Be the Key to Autonomous Cars

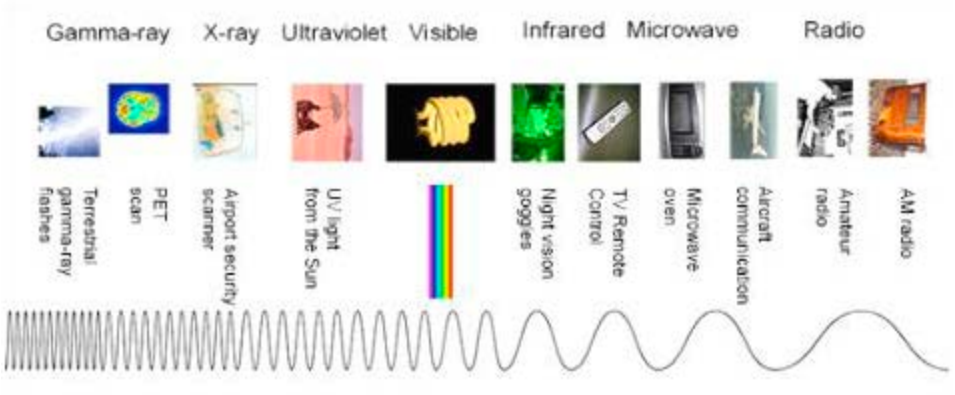
The New York Times

Sept. 26, 2019

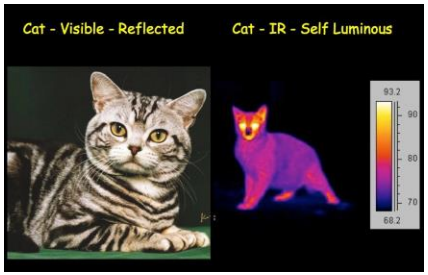
Human vision → Superhuman Vision

- Remote sensing use richer Sensors than AI (Computer Vision):
- Ex. Electromagnetic radiation Emitted by objects above absolute zero (0° Kelvin(K), −273°C)

Human



Source: imagine.gsfc.nasa.gov



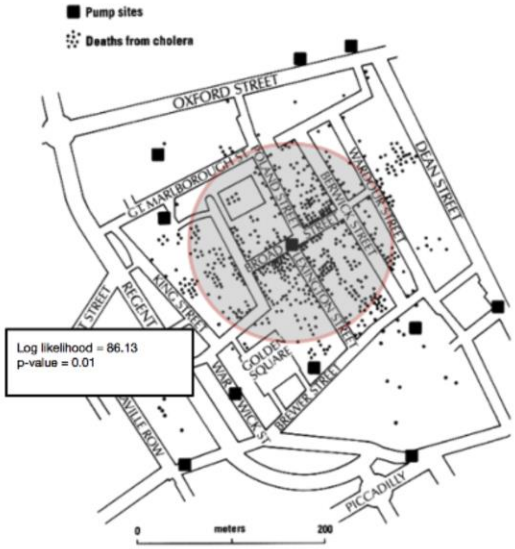
Source: directthermography.co.uk

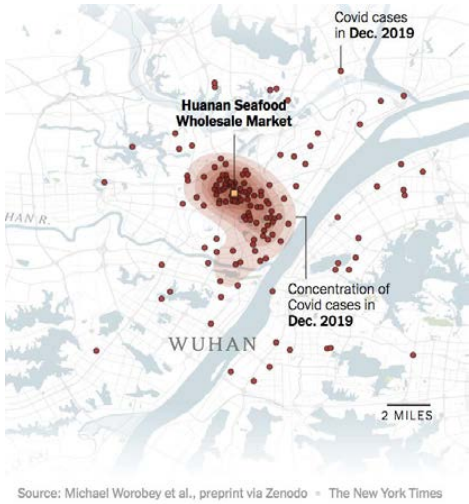
From Objects Detection to Pattern Mining

After object detection we look for Spatial patterns

- Hotspots, Spatial clusters
- Spatial outlier, discontinuities
- Co-locations, segregations
- Spatiotemporal predictions

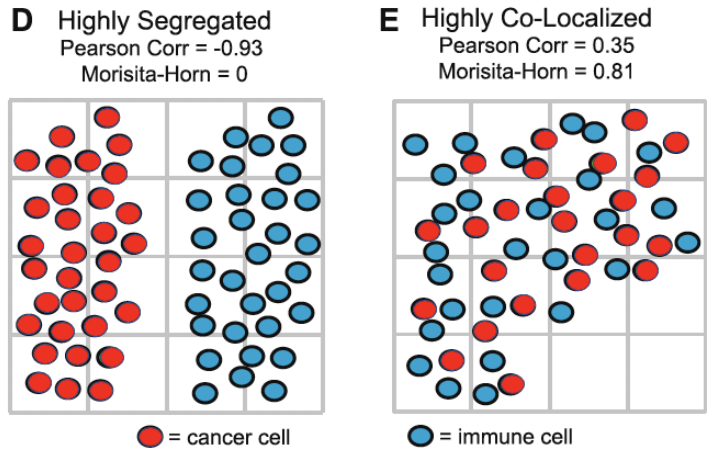
…





Spatial Data Mining is
The process of discovering

- interesting, useful, non-trivial patterns
- from large spatial datasets



Source: An ecological measure of immune-cancer colocalization as a prognostic factor for breast cancer. C. Maley et al., Breast Cancer Res 17, 131 (2015).

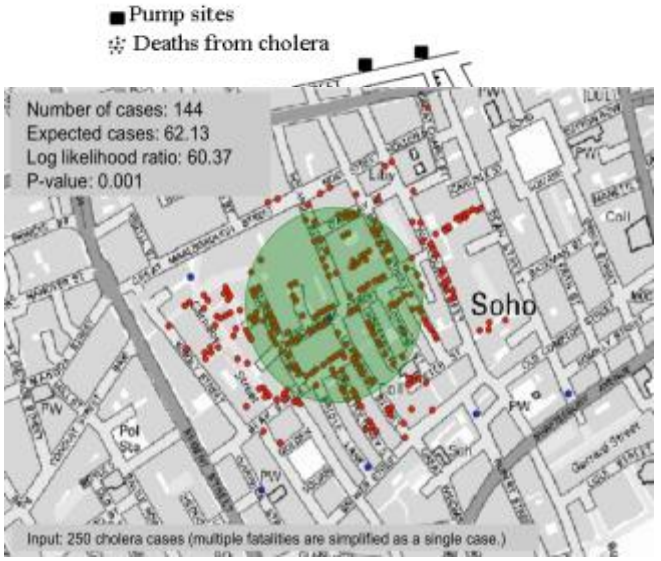
Spatial Pattern Mining Challenges

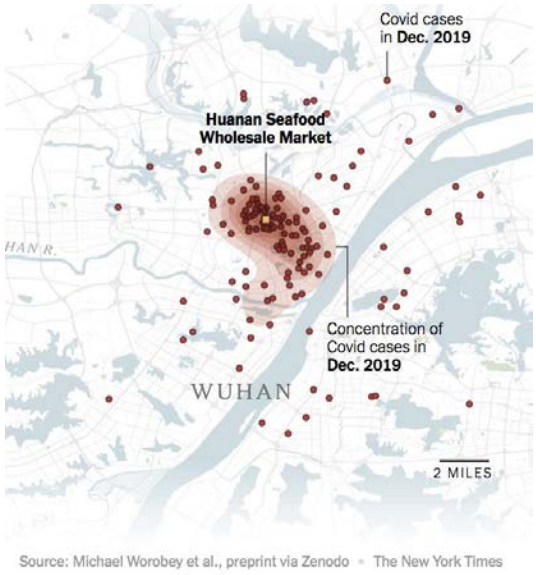
- Traditional pattern mining methods not robust in face of
  - Challenge 1: Noise
  - Challenge 2: Spatial continuity
  - Challenge 3: Auto-correlation, Heterogeneity , Edge-effect, …

Details: Data Science for Earth: The Earth Day Report , E. Eftelioglu, S. Shekhar, J. Hudson, L. Joppa, C. Baru, V. Janeja, et al. ACM SIGKDD Explorations Newsletter, 22(1), May 2020

Dealing with Noise & Chance Patterns

- Statistics: Deal with Noise
  - Quantify uncertainty, confidence, …
  - Is it (statistically) significant?
  - Is it different from a chance event or rest of dataset?
    - e.g., SaTScan finds circular hot-spots
- Spatial Statistics, Spatial Data Mining
  - Auto-correlation, Heterogeneity, Edge-effect, …

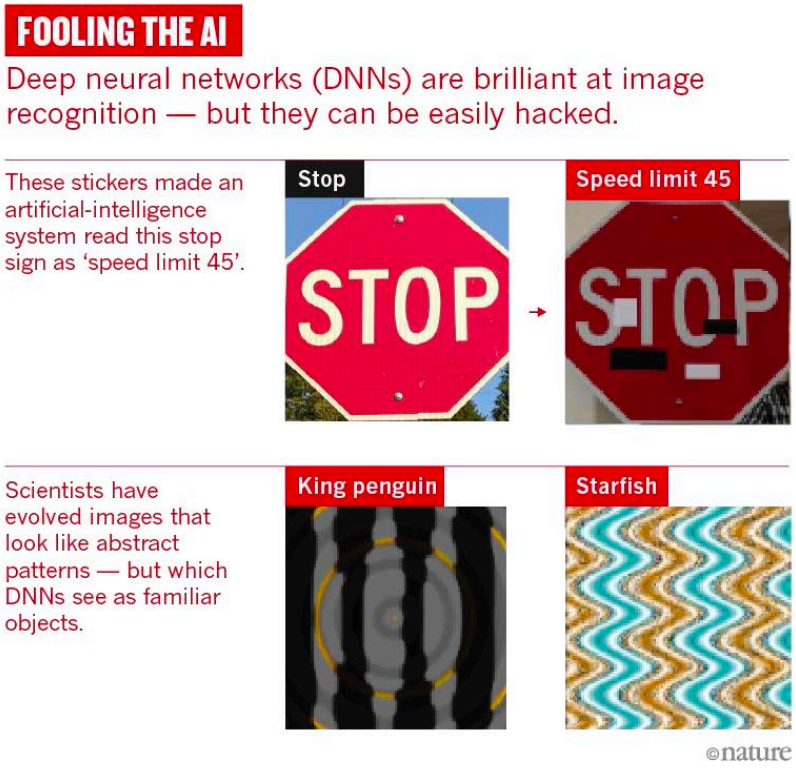


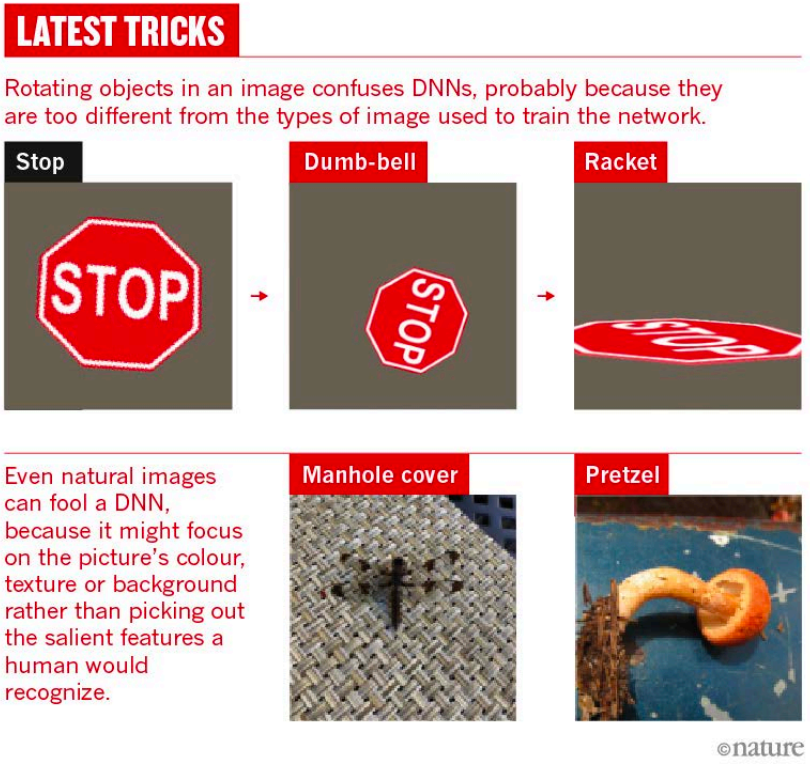


SaTScan™

Software for the spatial, temporal, and space-time scan statistics

However, machine is still learning

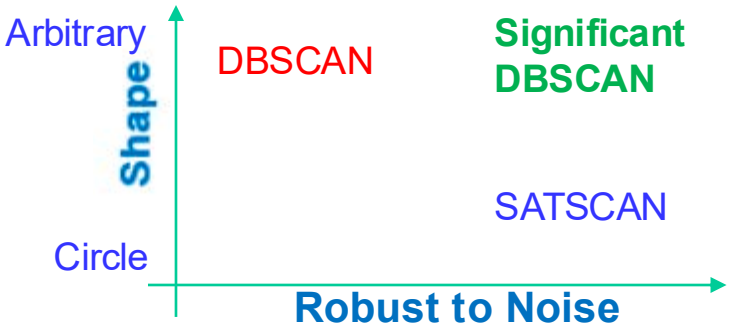




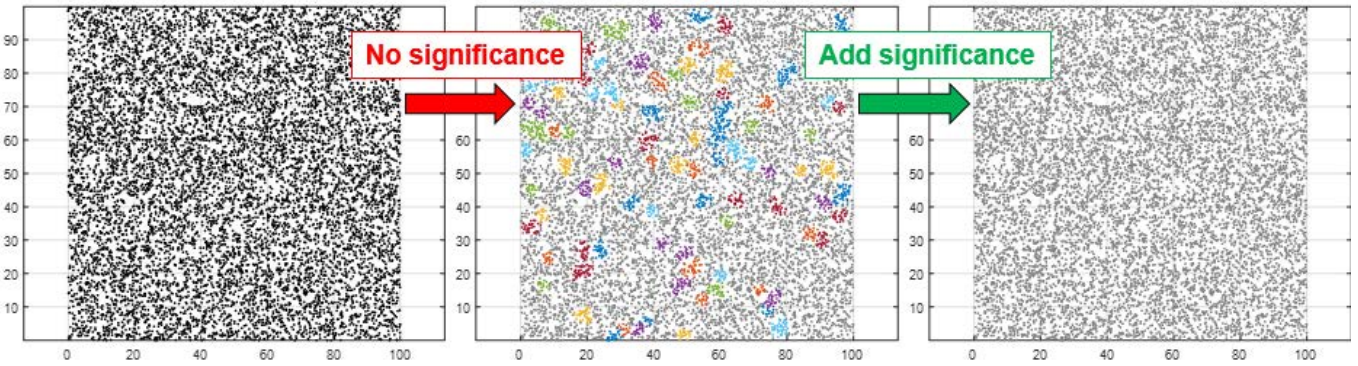
Source: [Why deep-learning AIs are so easy to fool](), D. Heaven, Nature, 09 Oct. 2019.

Robust Clustering (Hotspot Detection)

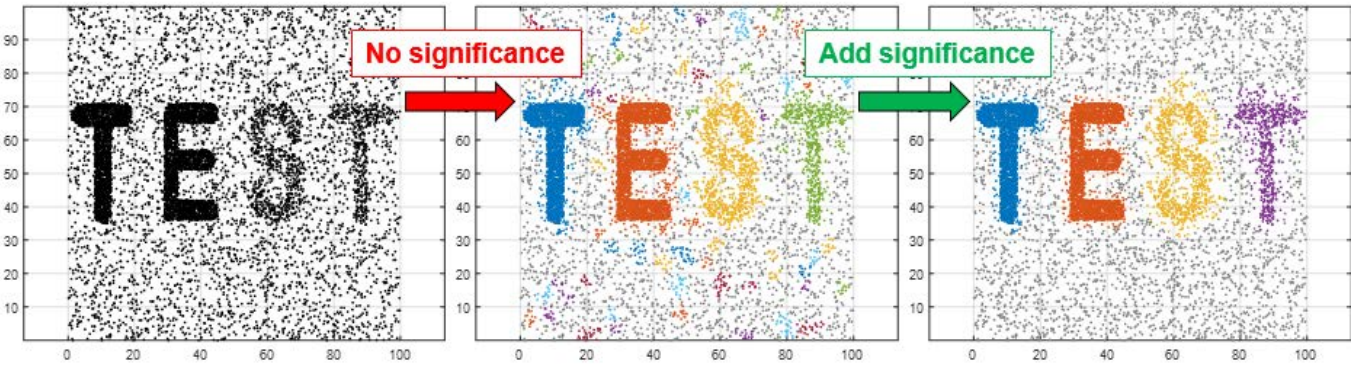
- **Problem definition**
  - **Inputs:** Collection of event locations, Test statistic; Significance level
  - **Output:** Significant clusters (hotspots)
  - **Constraints:** Avoid chance patterns despite non-trivial noise in data



Complete Spatial Random (noise)



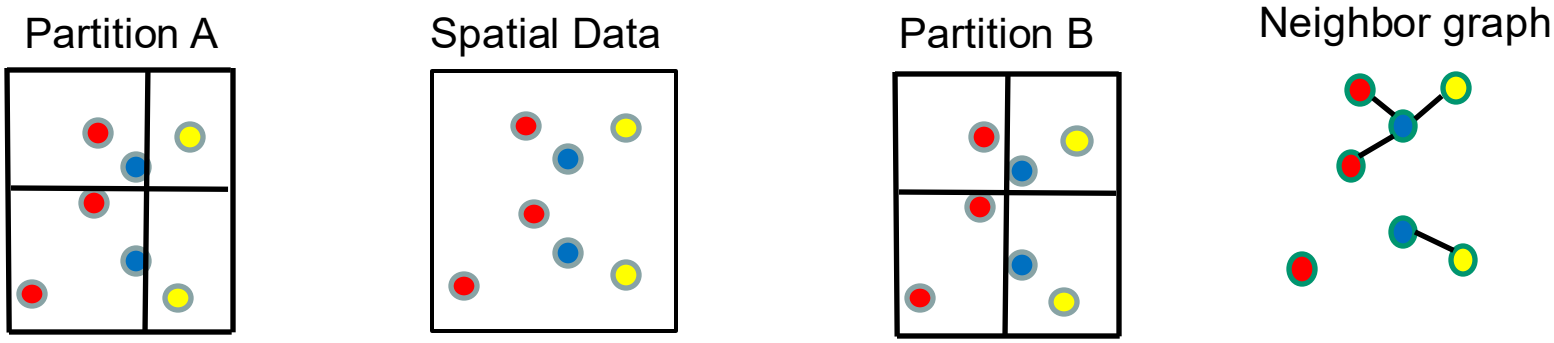
Hotspots with Noise



Details: Significant DBSCAN towards Statistically Robust Clustering, ACM Trans. on Intelligent Systems and Tech, 12(5):1-26, Oct. 2021. (A summary in 16th Intl. Symp. on Spatial and Temporal Databases, 2019. **(Best Paper Award)**

Challenge 2: Continuous Space

- Traditional relationship mining methods not robust
  - Result changes if spatial partitioning changes
  - Similar to Gerrymandering risk, Formally, Modifiable Areal Unit Problem (MAUP)
  - Neighbor Graph Based Measures are more robust



Partition A Based Pearson's Correlation	Pairs	Partition B Based Pearson's Correlation	Ripley's Cross-K	Participation Index
1	-	- 0.90	0.33	0.66
- 0.90	-	1	0.5	1

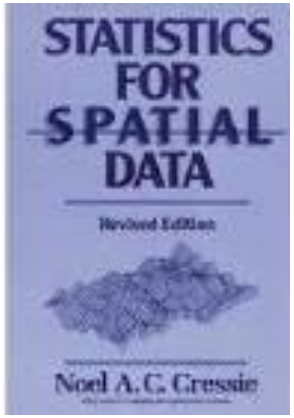
Details: Data Science for Earth: The Earth Day Report , E. Eftelioglu, S. Shekhar, J. Hudson, L. Joppa, C. Baru, V. Janeja, et al. ACM SIGKDD Explorations Newsletter, 22(1), May 2020

A Metric of Spatial Cross-Correlation

- Ripley's Cross K-Function Definition

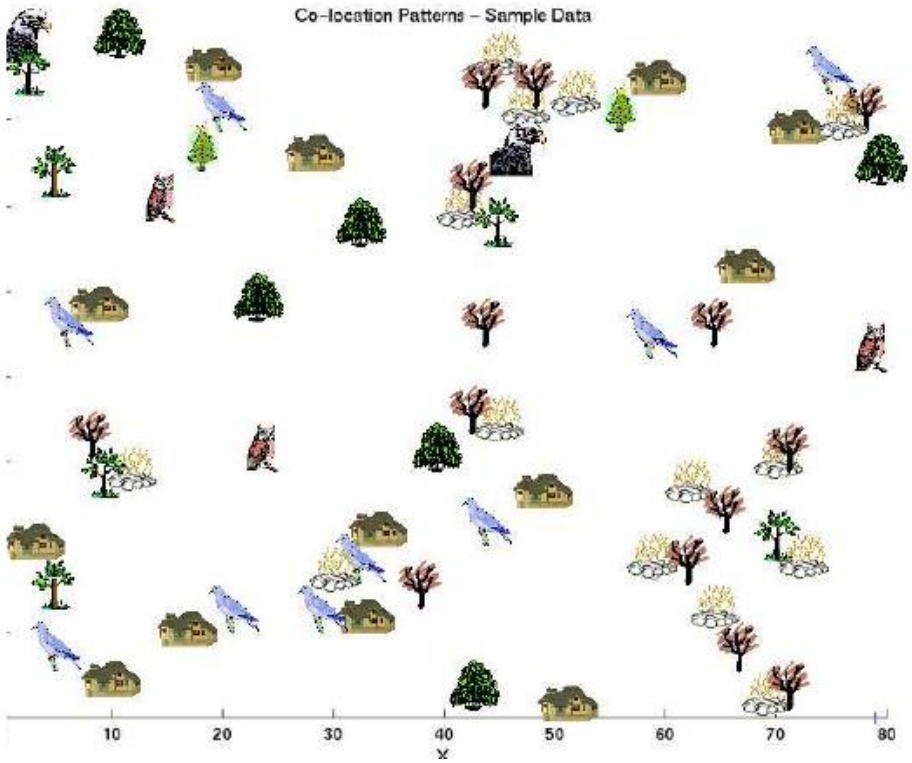
$$K_{ij}(h) = \lambda_j^{-1} E \left[\begin{array}{l}\text{number of type } \textbf{\textit{j}} \text{ event within distance } h \\ \text{of a randomly chosen type } \textbf{\textit{i}} \text{ event}\end{array}\right]$$

  - Interpretation
    - Which pairs are frequently co-located
    - Statistical significance



Co-locations

- **Given:** A collection of different types of spatial events
- **Find:** Co-located subsets of event types



Answers:   and  

Source: Discovering Spatial Co-location Patterns: A General Approach, IEEE Transactions on Knowledge and Data Eng., 16(12), December 2004 (w/ H.Yan, H.Xiong).

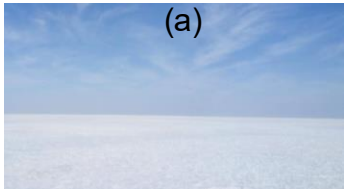
Modeling Spatial Auto-correlation

- Traditional Models, e.g., Regression (with Logit or Probit),
  - Linear Regression (LR), Bayes Classifier, …
- Semi-Spatial: LR with auto-corr. regularizer $\varepsilon = \|y - \beta X\|^2 + \left\|\beta X - \beta X_{neighbor}\right\|^2$
- Spatial
  - Spatial autoregressive model (SAR)
  - Markov random field (MRF) based Bayesian Classifier

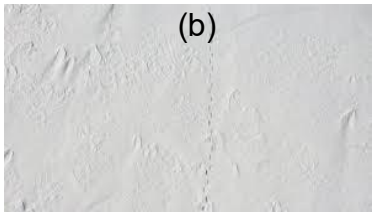
Traditional	Spatial
$y = X\beta + \varepsilon$	$y = \rho W y + X\beta + \varepsilon$
$\Pr(C_i \mid X) = \frac{\Pr(X \mid C_i)\Pr(C_i)}{\Pr(X)}$	$\Pr(c_i \mid X, C_N) = \frac{\Pr(C_i)\Pr(X, C_N \mid c_i)}{\Pr(X, C_N)}$
Neural Networks	Convolutional Neural Networks
Decision Trees	Spatial Decision Trees

Spatial Heterogeneity

- Knowledge of location can improve land-cover and object recognition
  - Q? Which pictures show snow?



Salt Marsh
(Runn of Kutch, Gujarat, India)



Snow

Snow

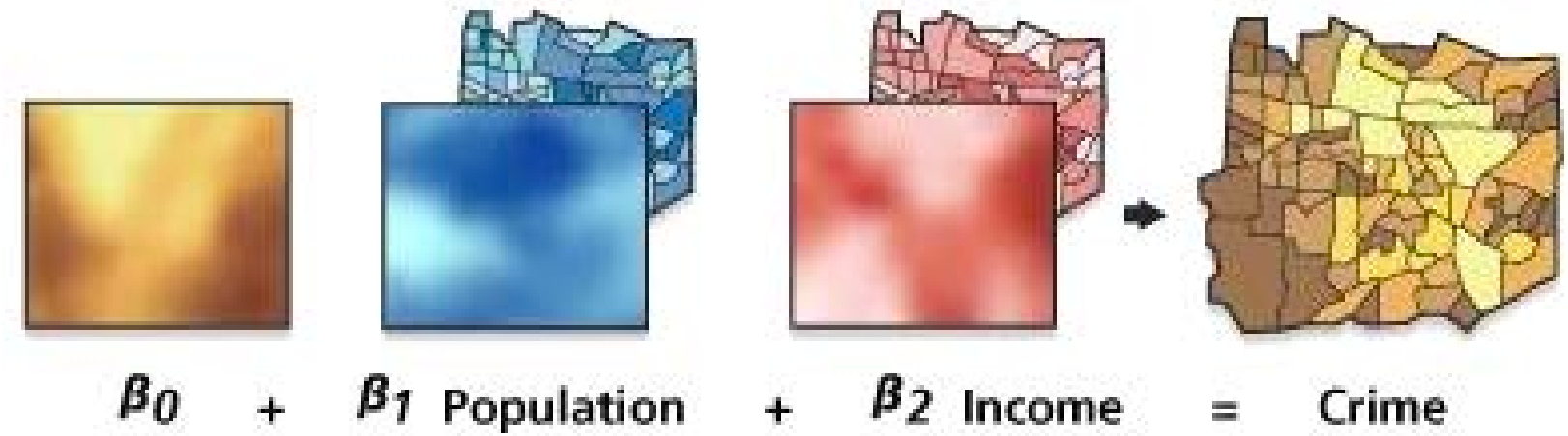
- Coarse Satellite Imagery (e.g., 30m pixels)
  - More effective for large mono-crop farms the small mixed-crop plots

- However, Convolutional Neural Networks does not model geographic heterogeneity.

Modeling Spatial Heterogeneity: GWR

- Geographically Weighted Regression (GWR)
  - Goal: Model spatially varying relationships
  - Example: $y = X\beta' + \varepsilon'$

    Where β' and ε' are location dependent

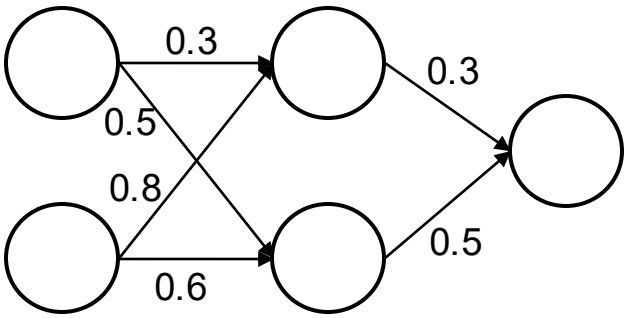


Source: resources.arcgis.com

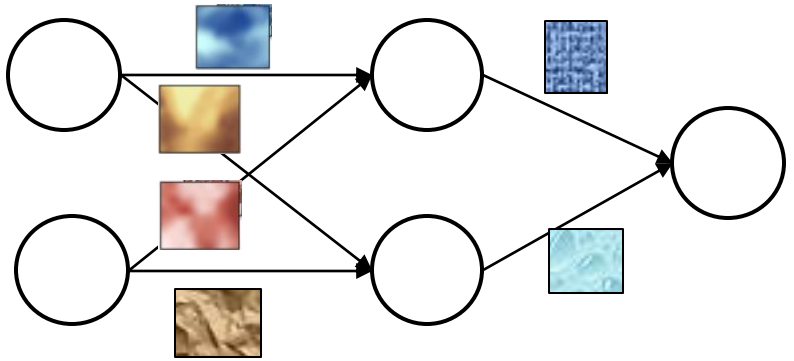
Spatial Variability Aware Neural Networks (SVANN)

- Each NN parameter is a map i.e., a function of location
  - Similar to Geographically Weighted Regression

A Neural Network (NN)



SVANN



- Evaluation:
  - Urban Garden Detection across Hennepin County, MN and Fulton County, GA.
  - SVANN outperformed OSFA by 14.34% on F1-scores.

Details: Towards Spatial Variability Aware Deep Neural Networks (SVANN), ACM Transactions on Intelligent Systems and Technology, 12(6):1-21, Dec. 2021. (A Summary in ACM SIGKDD Workshop on Deep Learning for Spatiotemporal Data, Applications, and Systems, 2020. (Best Paper Award)

A UCGIS Call to Action:
Bringing the Geospatial Perspective to Data Science Degrees and Curricula

Data that are geographically referenced or contain some type of location markers are both common and of high value (e.g., data subject to state-specific policies, laws and regulations; demographic data from the census; location traces of smartphones and vehicles; remotely sensed imagery from satellites, aircraft and small unmanned aerial vehicles; volunteered geographic information; geographically referenced social media postings). A 2011 McKinsey Global Institute report estimates a value of "about $600 billion annually by 2020" from leveraging personal location data[2] to reduce fuel waste, improve health outcomes, and better match products to consumer needs. Spatial data are critical for societal priorities such as national security, public health & safety, food, energy, water, smart cities, transportation, climate, weather, and the environment. For example, remotely-sensed satellite imagery is used to monitor not only weather and climate but also global crops[3] for early warnings and planning to avoid food shortages.

One Size Data Science Does not Fit All Data!

However, spatial data presents unique data science challenges. Recent court cases that address gerrymandering, the manipulation of geographic boundaries to favor a political party, offer a high-profile example. Instances of such exploitation of the modifiable areal unit problem (or dilemma) is not limited to elections since the MAUP affects almost all traditional data science methods in which results (e.g., correlations) change dramatically by varying geographic boundaries of spatial partitions. The fundamental geographic qualities of spatial autocorrelation, which assumes properties of geographically proximate places to be similar, and geographic heterogeneity, where no two places on Earth are exactly alike, violate assumptions of sample independence and randomness that underlie many conventional statistical methods. Other spatial challenges include how to choose between a plurality of projections and coordinate systems and how to deal with the imprecision, inaccuracy, and uncertainty of location

A UCGIS Call to Action:

Bringing the Geospatial Perspective to Data Science Degrees and Curricula

Spatial Data Science Tools

measurements. To deal with such challenges, practitioners in many fields including agriculture, weather forecast, mining, and environmental science incorporate *geospatial data science*[4] methods such as spatially-explicit models, spatial statistics[5], geo-statistics, geographic data mining[6], spatial databases[7], etc.

[4] Y. Xie et al., Transdisciplinary Foundations of Geospatial Data Science, *ISPRS Intl. Jr. of Geo-Informatics*, 6(12):395-418, 2017. DOI: 10.3390/ijgi6120395.

[5] N. Cressie, *Statistics for Spatial Data*, Wiley, 1993 (1st ed.), 2015 (Revised ed.).

[6] H. Miller and J. Han, *Geographic Data Mining and Knowledge Discovery*, CRC Press, 2009 (2nd Ed.).

[7] S. Shekhar and S. Chawla, *Spatial Databases: A Tour*, Prentice Hall, 2003.

Summary

- GeoAI has already transformed our society
  - It is only a beginning!
  - It promises astonishing opportunities

- GeoAI has promise but faces major challenges
  - High cost of errors, Spatial Heterogeneity, …

- **Ask**
  - Sponsors: Nurture approaches to overcome challenges
  - Academics: Include Spatial topics in courses and curricula

University Consortium for GEOGRAPHIC INFORMATION SCIENCE
Summer 2018

A UCGIS Call to Action:
Bringing the Geospatial Perspective to Data Science Degrees and Curricula

References :Surveys, Overviews

- Spatial Computing, MIT Press (Essential Knowledge Series), 2020. (ISBN: 9780262538046).
- Spatial Computing, Communications of the ACM, 59(1):72-81, January, 2016.
- Transdisciplinary Foundations of Geospatial Data Science, ISPRS Intl. Jr. of Geo-Informatics, 6(12):395-429, 2017. (doi:10.3390/ijgi6120395)
- Spatiotemporal Data Mining: A Computational Perspective, ISPRS Intl. Jr. on Geo-Information, 4(4):2306-2338, 2015 (DOI: 10.3390/ijgi4042306).
- Identifying patterns in spatial information: a survey of methods, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1(3):193-214, May/June 2011.
- Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data, IEEE Transactions on Knowledge and Dat Mining, 29(10):2318-2331, June 2017.
- Parallel Processing over Spatial-Temporal Datasets from Geo, Bio, Climate and Social Science Communities: A Research Roadmap. IEEE BigData Congress 2017: 232-250.
- Spatial Databases: Accomplishments and Research Needs, IEEE Transactions on Knowledge and Data Engineering, 11(1):45-55, 1999.