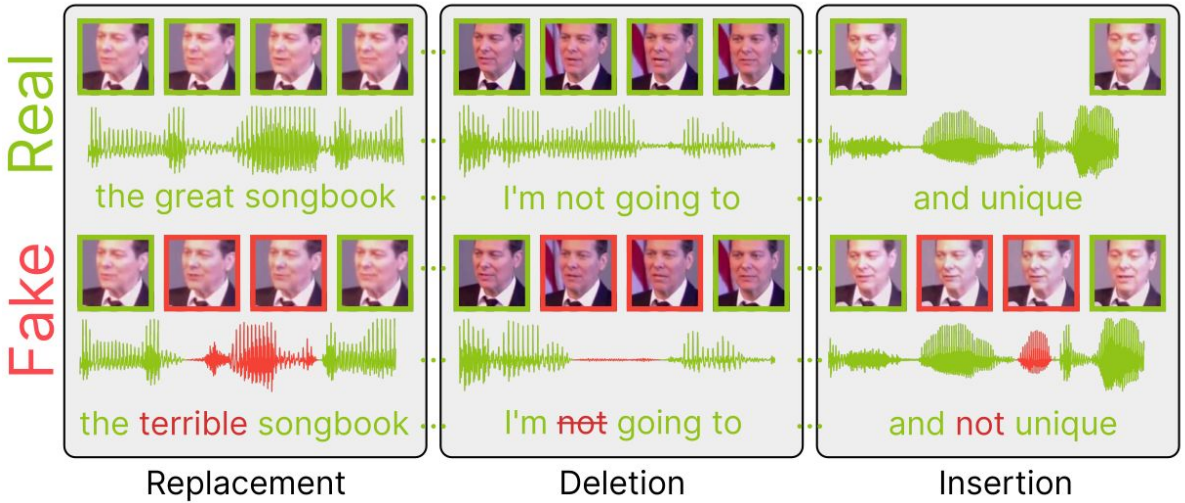
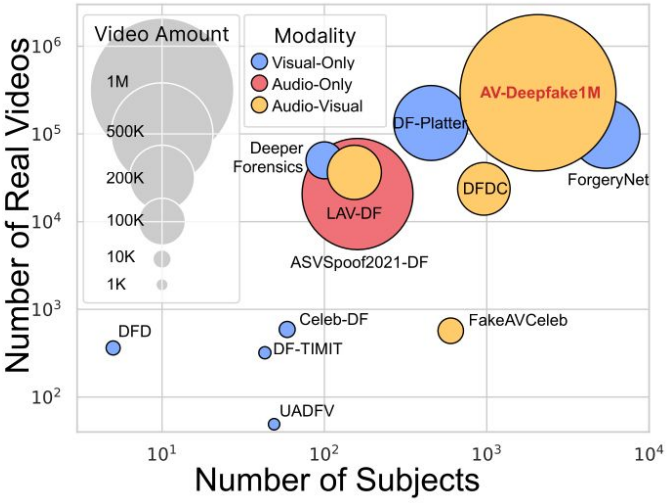
Thematic session

Paper presentation: Group3

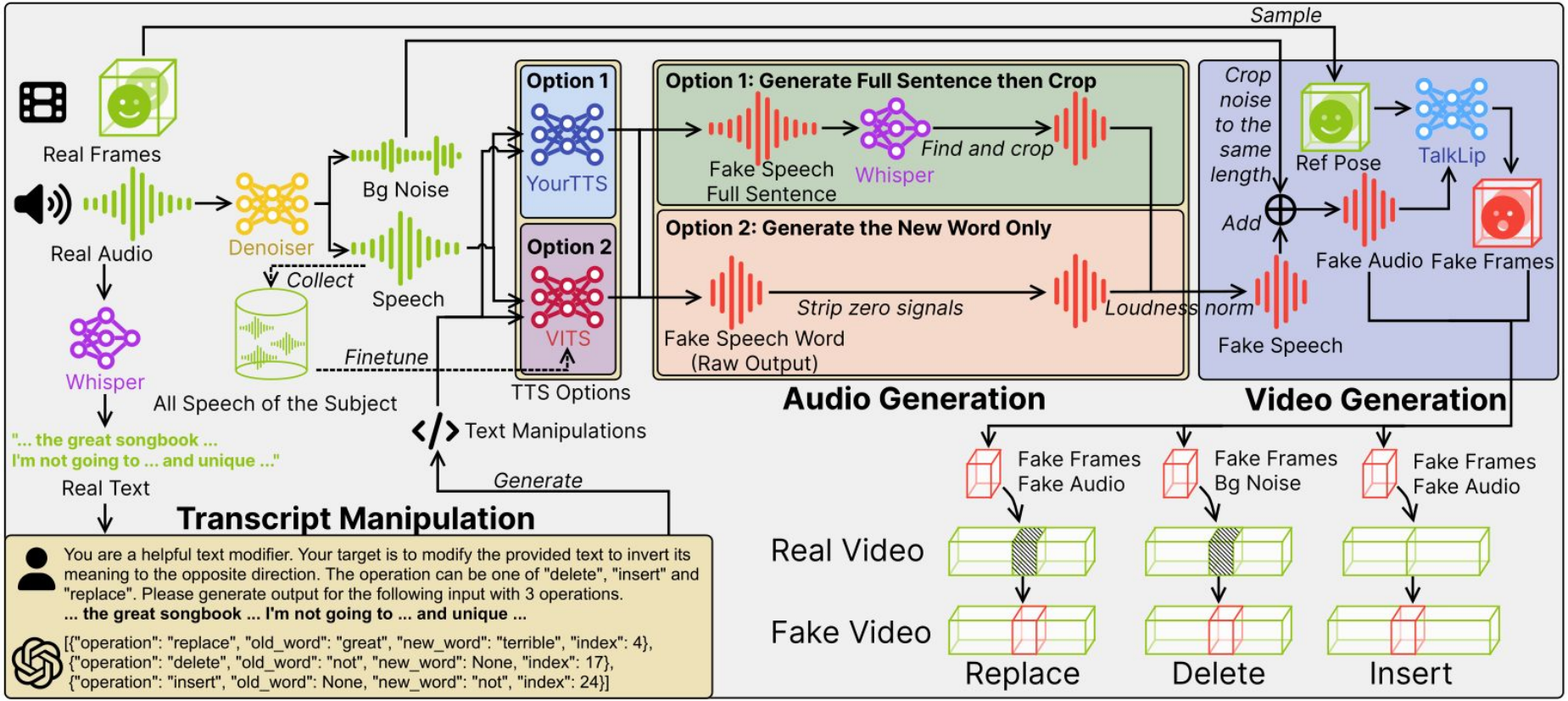
Stevan Rudinac





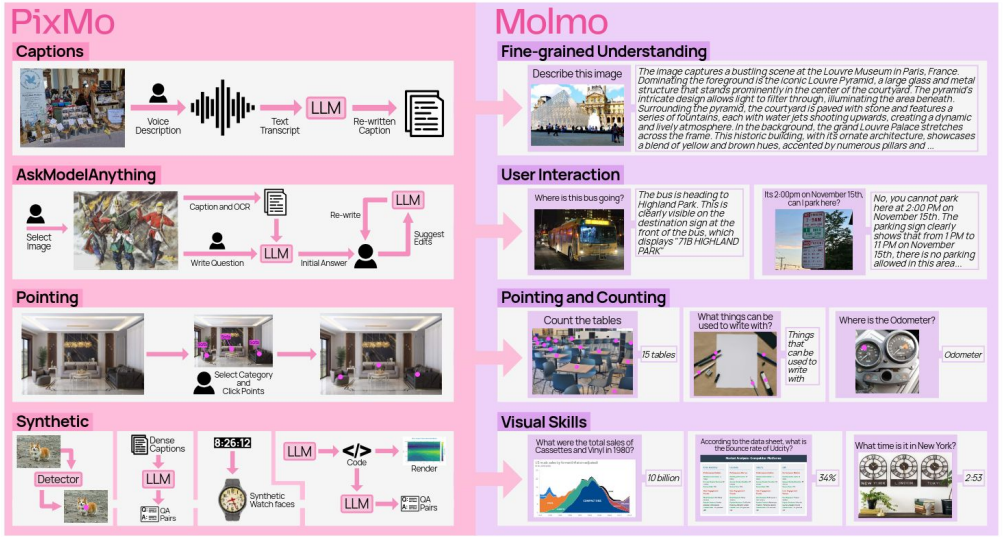
- AV-Deepfake1M is a large-scale content-driven deepfake dataset generated using a large language model.
- Best Student Paper Award at ACM Multimedia 2024 in Melbourne.

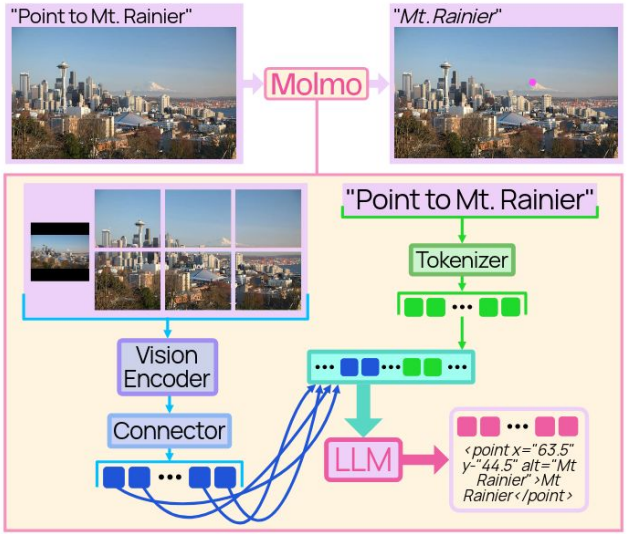
Zhixi Cai, Shreya Ghosh, Aman Pankaj Adatia, Munawar Hayat, Abhinav Dhall, Tom Gedeon, and Kalin Stefanov. 2024. AV-Deepfake1M: A Large-Scale LLM-Driven Audio-Visual Deepfake Dataset. In Proceedings of the 32nd ACM International Conference on Multimedia (MM '24). Association for Computing Machinery, New York, NY, USA, 7414–7423. https://doi.org/10.1145/3664647.3680795



- **Preprocessing**: Audio extraction via FFmpeg followed by Whisper-based transcript generation.
- **Step 1 (transcript manipulation)**: The transcript is modified through word-level insertions, deletions & replacements.
- **Step 2 (audio generation)**: The audio is generated in both speaker-dependent and independent fashion.
- **Step 3 (video generation)**: Based on the generated audio, the subject-dependant video is generated with smooth transitions in terms of lip-synchronization, pose, and other relevant attributes.

Takeaway from Andrei Bursuc's MMM'25 Keynote

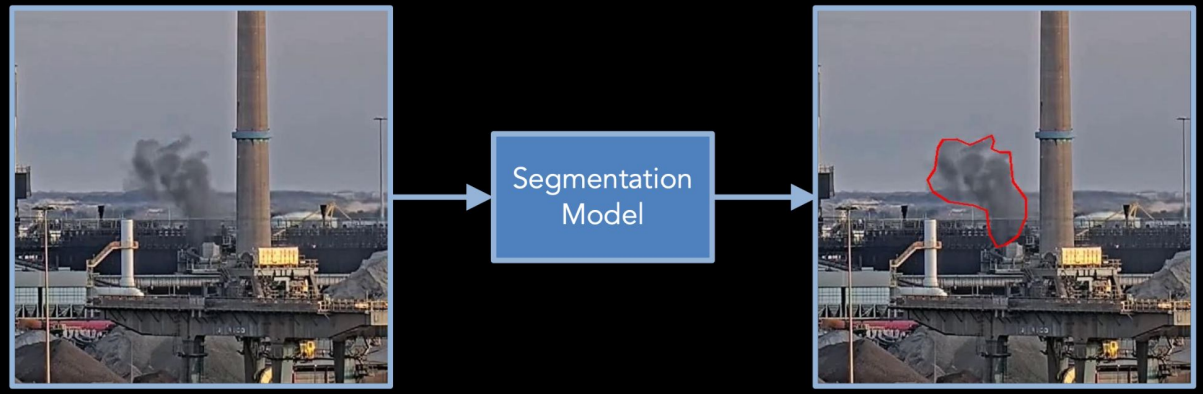




- Careful and smart data selection and annotation can go a long way
- Molmo is a very competitive VLM from Ai2, trained on 700k image/caption pairs
- 3 annotations per image; annotation speech is recorded for 60-90 seconds; formatted questions

Deitke, M., Clark, C., Lee, S., Tripathi, R., Yang, Y., Park, J. S., ... & Kembhavi, A. (2024). Molmo and pixmo: Open weights and open data for state-of-the-art multimodal models. *arXiv preprint arXiv:2409.17146*. (https://molmo.allenai.org/)

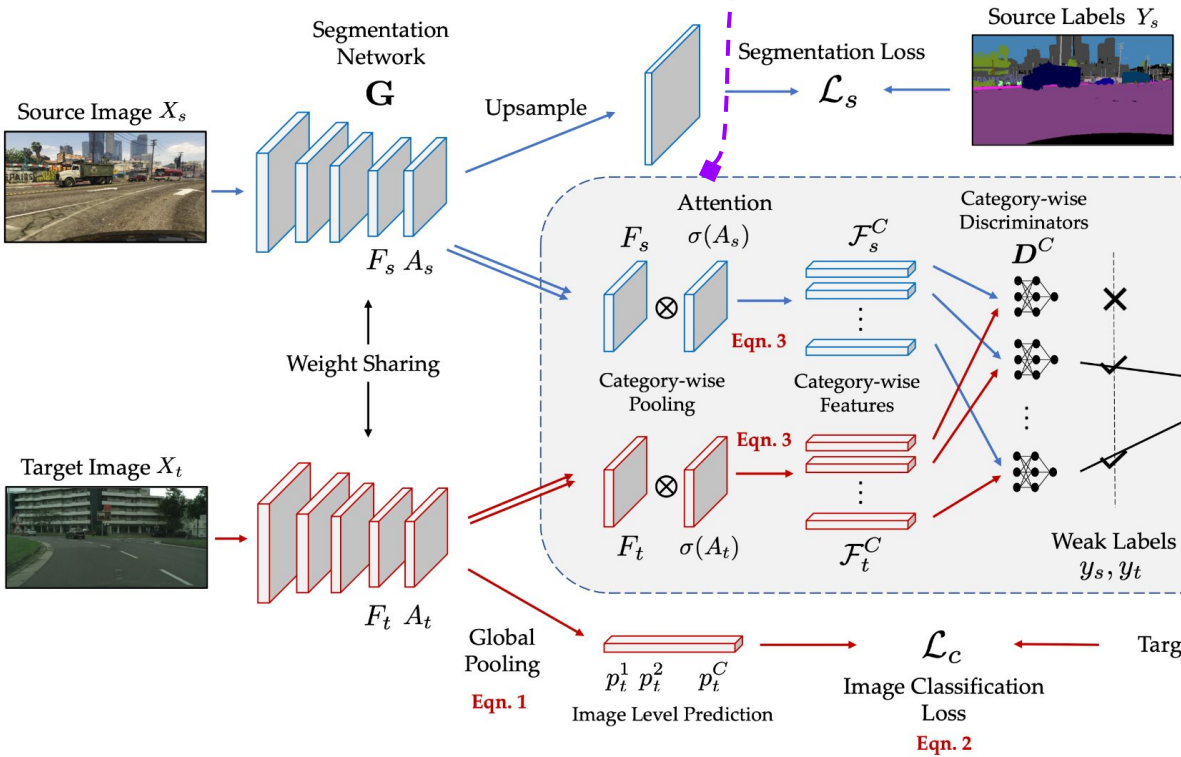
Yen-Chia Hsu



To what extent can community feedback help improve the model?

Wrong: No Smoke | Wrong: Steam | Too Small | Roughly OK

Yen-Chia Hsu



Why using attention and not just max/average pooling or a simple linear layer?

What if we also have some segmentation labels from the target domain? What if we have no access to source labels but have the model?

How to do this when we have bounding box instead? What if we have more types of weak labels?

How to replace this with bounding box? Change the loss function?

Domain Adaptive Semantic Segmentation Using Weak Labels, ECCV 2020, https://arxiv.org/abs/2007.15176

Tim Alpherts

Detecting disparities in police deployments using dashcam data

Matt Franchi
mwf62@cornell.edu
Cornell Tech
New York, New York, USA

J.D. Zamfirescu-Pereira
University of California - Berkeley
Berkeley, USA
zamfi@berkeley.edu

Wendy Ju
Jacobs Technion-Cornell Institute, Cornell Tech
New York, USA

Emma Pierson
Jacobs Technion-Cornell Institute, Cornell Tech
New York, USA

Problem Statement

- Policing data is vital for detecting inequity in police behavior
- Aggregated police deployment data is unavailable

Problem Statement

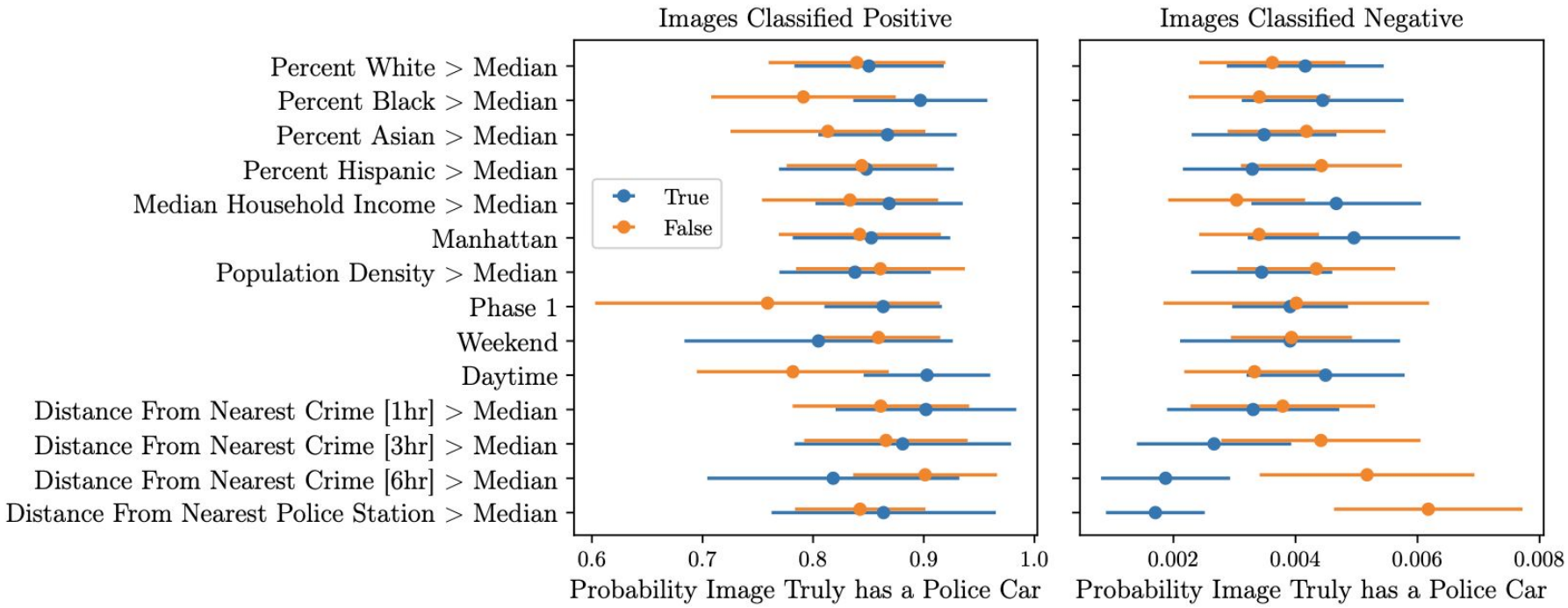
- Downstream policing data has only been previously studied
- Studies are qualitative
- Existing quantitative studies use datasets provided by the police themselves
- Predictive policing data depends on downstream data
- If police deployment is biased, this could affect downstream data and incur model bias

Method

- Dataset by Nexar
- 24,803,854 images taken throughout the five boroughs of New York City between March 4 2020 and November 15 2020. Each image is 1280 x 720 pixels.
- Label 15000 images for police vehicles (Sedans, SUVs, Compacts, and Trucks)
- Train with YOLO

(a) True Positives

(b) False Positives


Images Classified Positive
Images Classified Negative
Percent White > Median
Percent Black > Median
Percent Asian > Median
Percent Hispanic > Median
Median Household Income > Median
Manhattan
Population Density > Median
Phase 1
Weekend
Daytime
Distance From Nearest Crime [1hr] > Median
Distance From Nearest Crime [3hr] > Median
Distance From Nearest Crime [6hr] > Median
Distance From Nearest Police Station > Median
True
False
0.6
0.7
0.8
0.9
1.0
0.002
0.004
0.006
0.008
Probability Image Truly has a Police Car
Probability Image Truly has a Police Car

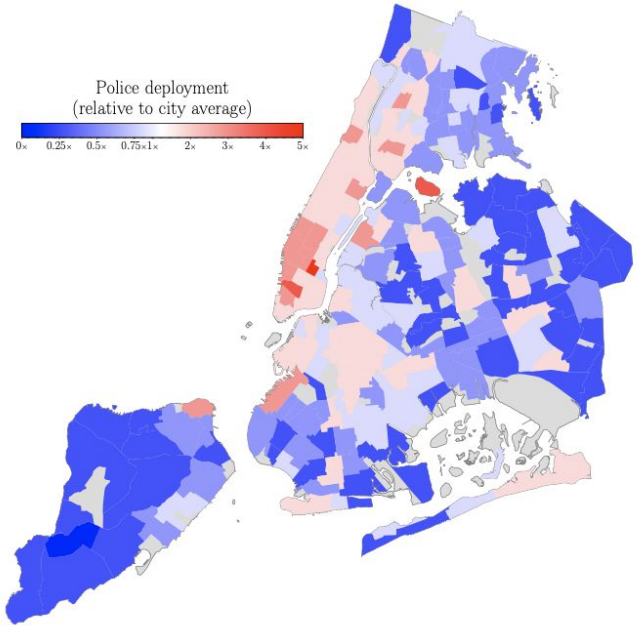
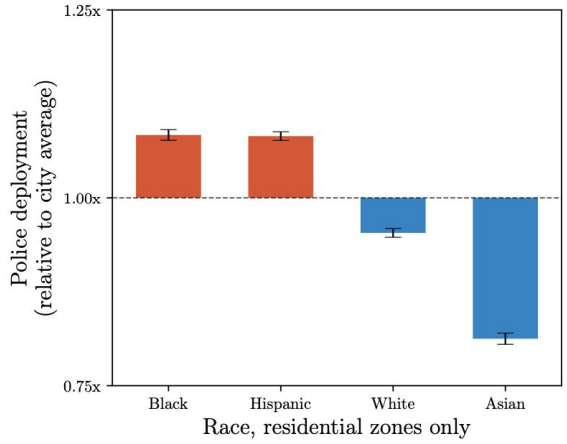
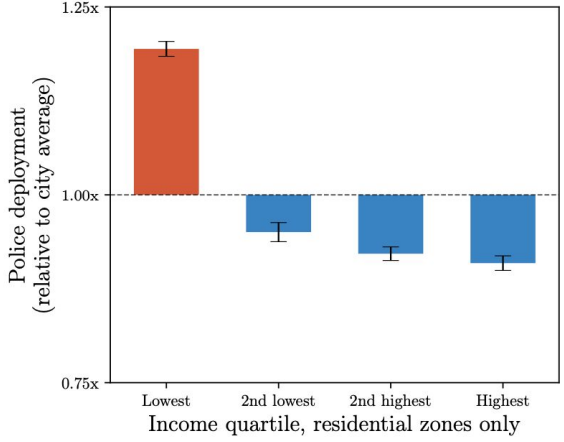


Figure 6: Map of police deployment throughout NYC, expressed relative to the city average. Grey areas are those with zero population in Census data, including airports, cemeteries, and parks.

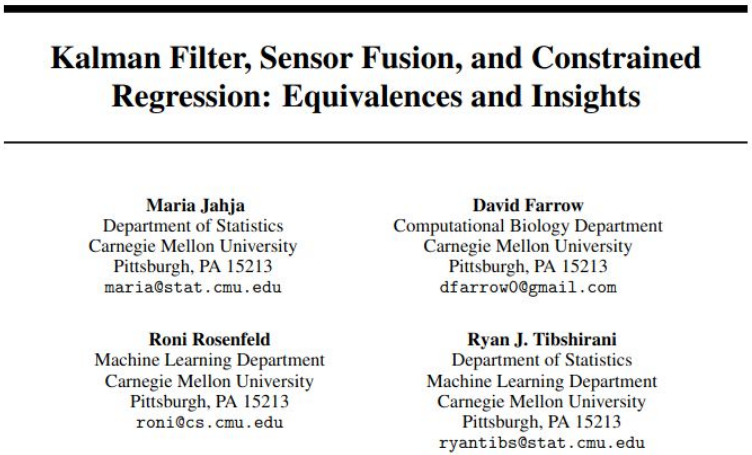


(a) Deployments by race group.



(b) Deployments by income quartile.

Yahia Dalbah



Kalman Filter, Sensor Fusion, and Constrained Regression: Equivalences and Insights

Maria Jahja
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
maria@stat.cmu.edu

David Farrow
Computational Biology Department
Carnegie Mellon University
Pittsburgh, PA 15213
dfarrow0@gmail.com

Roni Rosenfeld
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213
roni@cs.cmu.edu

Ryan J. Tibshirani
Department of Statistics
Machine Learning Department
Carnegie Mellon University
Pittsburgh, PA 15213
ryantibs@stat.cmu.edu

Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019

Short intro (1) Kalman Filters (KF)

Time-invariant linear dynamical systems (t = 1, 2, 3 …)

Actual States (process model, assumed) $x_t = Fx_{t-1} + \delta_t,$

Measured values (measurement model) $z_t = Hx_t + \epsilon_t,$

Actual state values (i.e. ground truth) is unobservable

Kalman filters (KF):

$\bar{x}_{t+1} = F\hat{x}_t,$ Intermediate estimate (predict step), use model and previous estimate

$\hat{x}_{t+1} = \bar{x}_{t+1} + K_{t+1}(z_{t+1} - H\bar{x}_{t+1}),$ Intermediate estimate

Remark:
If the process noises are gaussian, then KF is a discretized bayes estimator

Short intro (2) Sensor Fusion and Extended Kalman Filter

General case of KF (KF is SF when the system is already linear)
Aim is to linearize non-linear systems

Assume:
non-linear process model and maps

$$\bar{x}_{t+1} = f(\hat{x}_t),$$

$$H_{t+1} = Dh(\bar{x}_{t+1}) \qquad F_{t+1} = Df(\hat{x}_t)$$

Jacobian (df/dx, dh/dx) at (t-1)

Yields extended sensor fusion (ESF)/extended kalman filter (EKF)

Paper contribution

Edit the formulation of the problem such that the actual process value are observed with a time lag.

Example problem: Influenza (flu) nowcasting
Goal: Estimate the weekly percentage of weighted influenza-like illness (wILI)

Sensors (proxies) - United States:
Google Flu Trends and Google Health Trends
Google Trends (search terms)
Health Tweets
Electronic Health Records

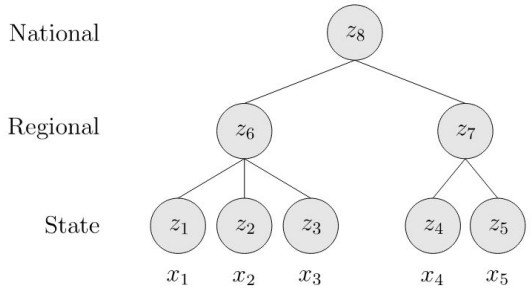
Webpage visits:
- Wikipedia
- Centers for Disease Control and Prevention

Target: Reduce the collection of many inaccurate and noisy sensors to a converging robust prediction

Problem formulation

- Replace noise covariance ‘ R ‘ by empirical covariance from past flu data
- Reduce the EKF estimate equation to a regression of states, where H would be reduced to regression coefficients

$$\hat{R}_{t+1} = \frac{1}{t}\sum_{i=1}^{t}(z_i - Hx_i)(z_i - Hx_i)^T,$$

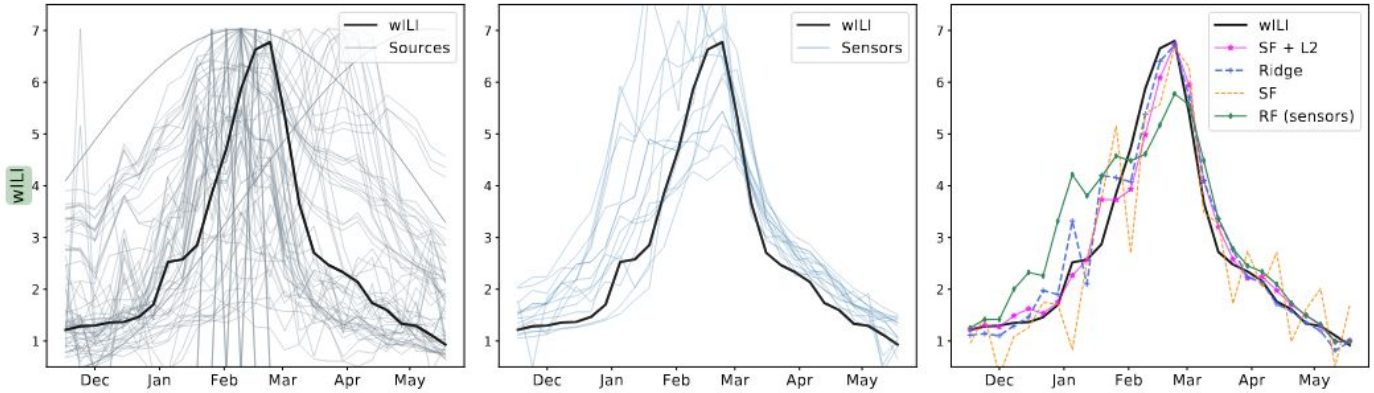


$$H = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1/3 & 1/3 & 1/3 & 0 & 0 \\ 0 & 0 & 0 & 1/2 & 1/2 \\ 1/5 & 1/5 & 1/5 & 1/5 & 1/5 \end{bmatrix}.$$

Instead of dumping the whole data as multi-input => output forecasting problem, transform it into multi-input => multi-weak regressions => output

Setup and results

- At week t+1, estimate $x\hat{}_{t+1}$ from a collection of ‘sensors (308 measurement proxies)
- Each proxy is a (weak) regression trained on t-155 weeks



Ujjwal Sharma