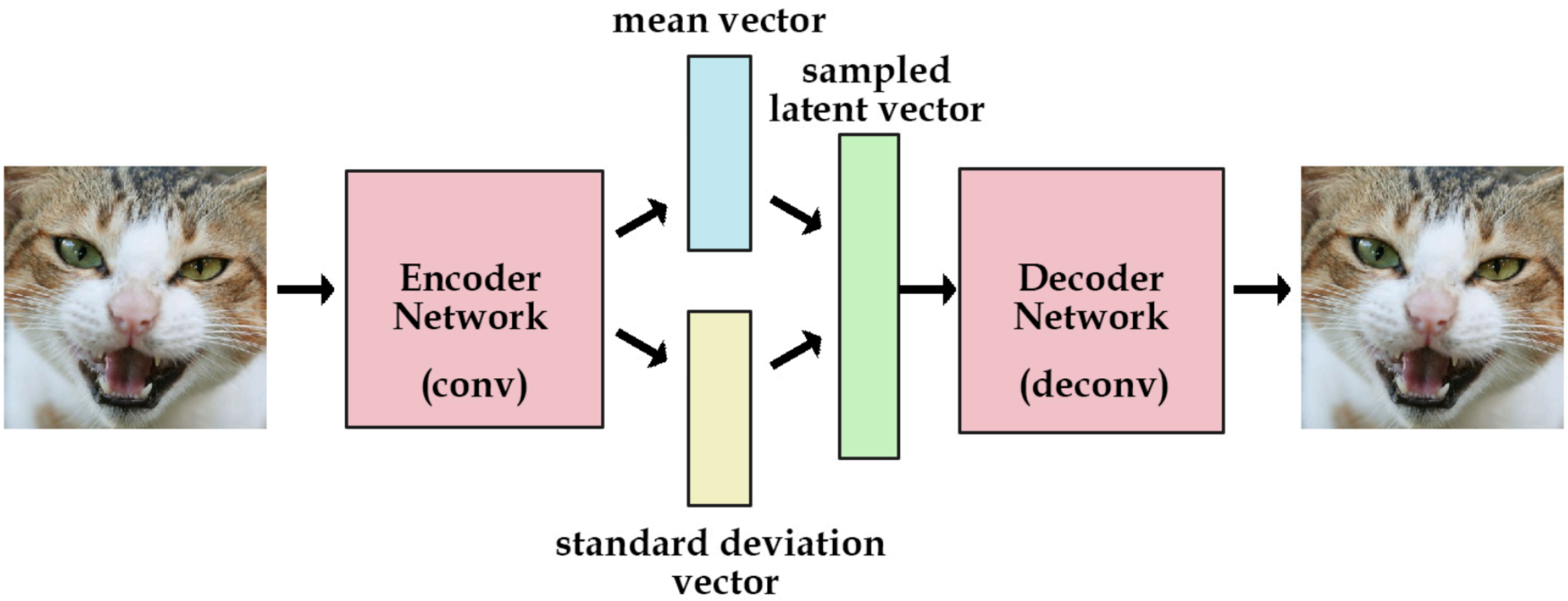
Unsupervised Anomaly Detection via Variational Auto-Encoder for Seasonal KPIs in Web Applications

By Haowen Xu et al.

aka the Donut paper

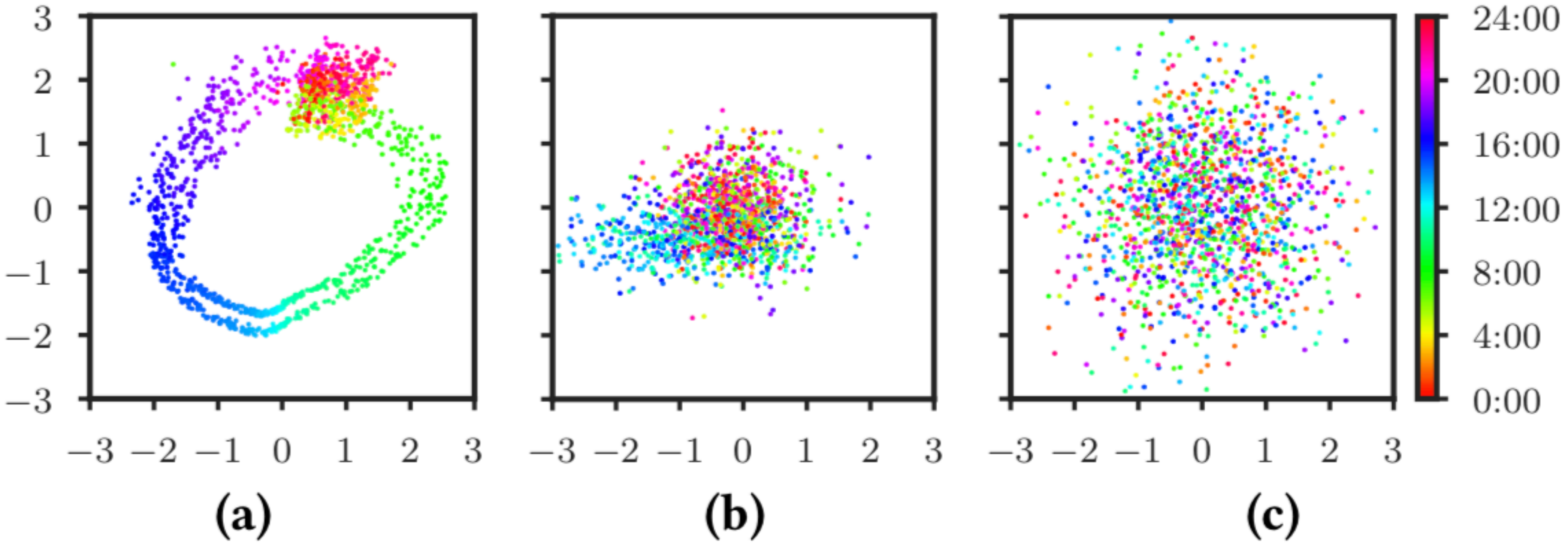
Key Ingredients for the Donut

- 🍩 Variational Auto-Encoder



Variational Autoencoders Visualized

Source: http://kvfrans.com/variational-autoencoders-explained/



Why Donut?

MATHEMATICIANS FARE POORLY IN WAR.
WHAT'S THAT SYMBOL MEAN?
NOTHING, BUT IF I CAN GET ENOUGH ARROW STRIKES, I CAN CALCULATE PI!

Monte Carlo Integration

Source: https://www.smbc-comics.com/comic/math-and-war

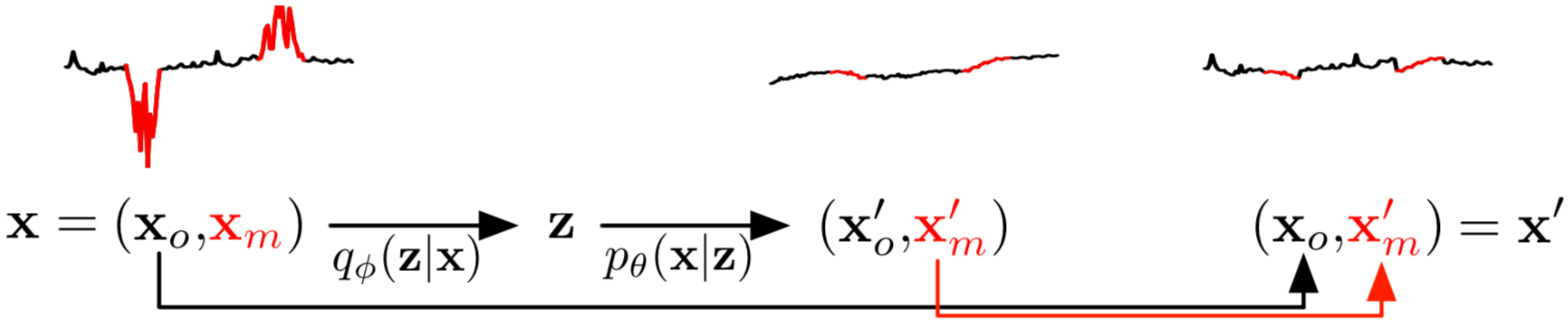
Key Ingredients for the Donut

- Variational Auto-Encoder
- Modified Evidence Lower Bound

“This modification trains Donut to correctly reconstruct the normal points within x, even if some points in x are abnormal.”

Key Ingredients for the Donut

- Variational Auto-Encoder
- Modified Evidence Lower Bound
- Missing Data Injection
- MCMC Imputation

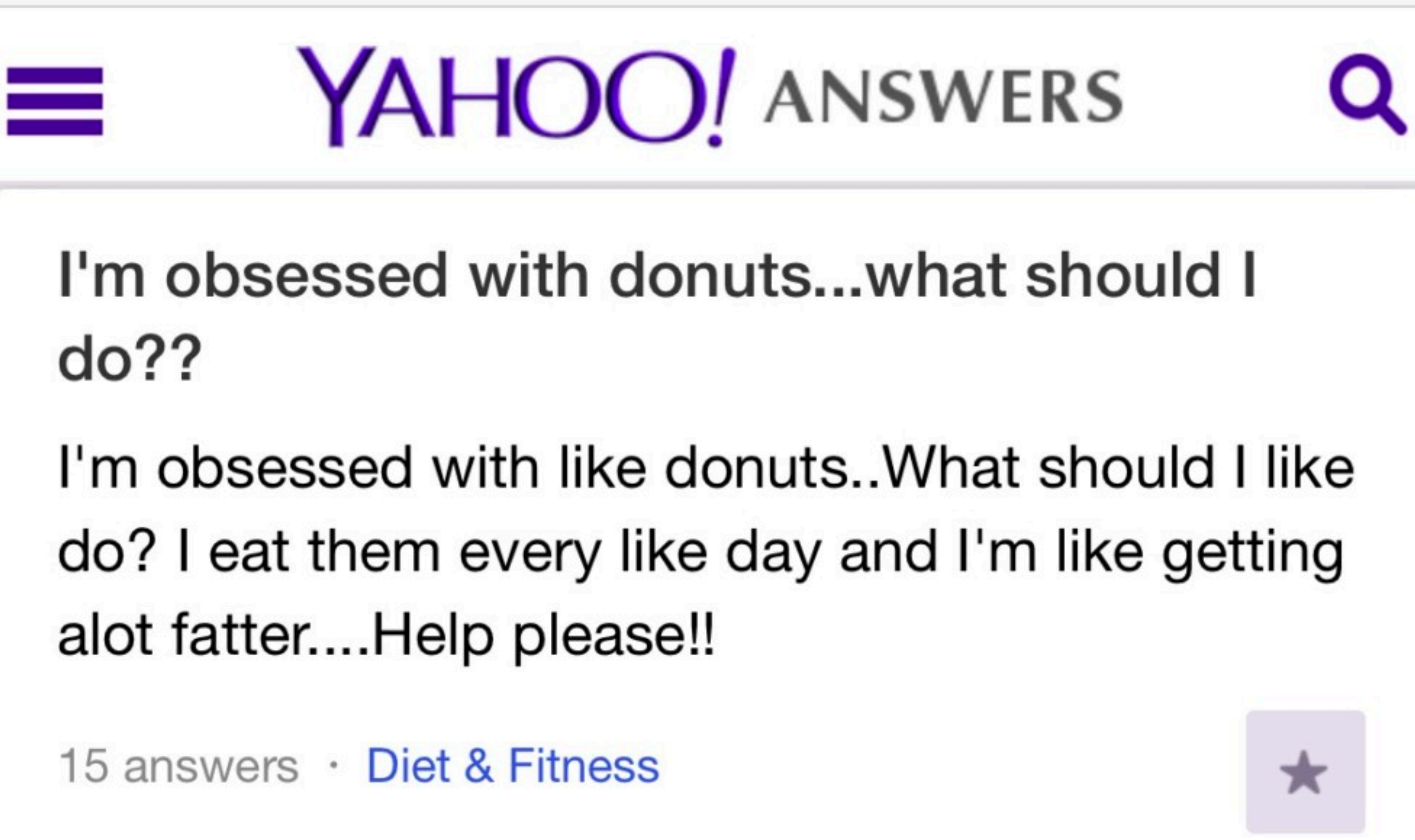


$$\mathbf{x} = (\mathbf{x}_o, \mathbf{x}_m) \xrightarrow[q_\phi(\mathbf{z}|\mathbf{x})]{} \mathbf{z} \xrightarrow[p_\theta(\mathbf{x}|\mathbf{z})]{} (\mathbf{x}'_o, \mathbf{x}'_m) \qquad (\mathbf{x}_o, \mathbf{x}'_m) = \mathbf{x}'$$

MCMC Imputation Single Pass

We build it not because we can but because we want to.

“…there is no theoretical foundation to back up its designs of deep generative models for anomaly detection”



I'm obsessed with donuts...what should I do??

I'm obsessed with like donuts..What should I like do? I eat them every like day and I'm like getting alot fatter....Help please!!

15 answers · Diet & Fitness

Yahoo!

Extensible Generic Anomaly Detection System aka EGADS

Source: http://blogdailyherald.com/2016/10/20/donut-get-food-tattoo/

EGADS

- Distinctions between outliers, anomalies, and change points
- Outlier detection is similar in definition to Donut’s goal
- Time-series modeling module also compensates for missing data points
- Relies on relative error for triggering anomalies
- Built for “Yahoo scale”
- Practical extensions around alerting

Themselves

Opprentice: Sorry past me

Source: thecoolhunter.net

Adult supervision required

- Primary stated goal was to simplify use/deployment of anomaly detectors

- Used a lot of “basic anomaly detectors” with various configurations as features

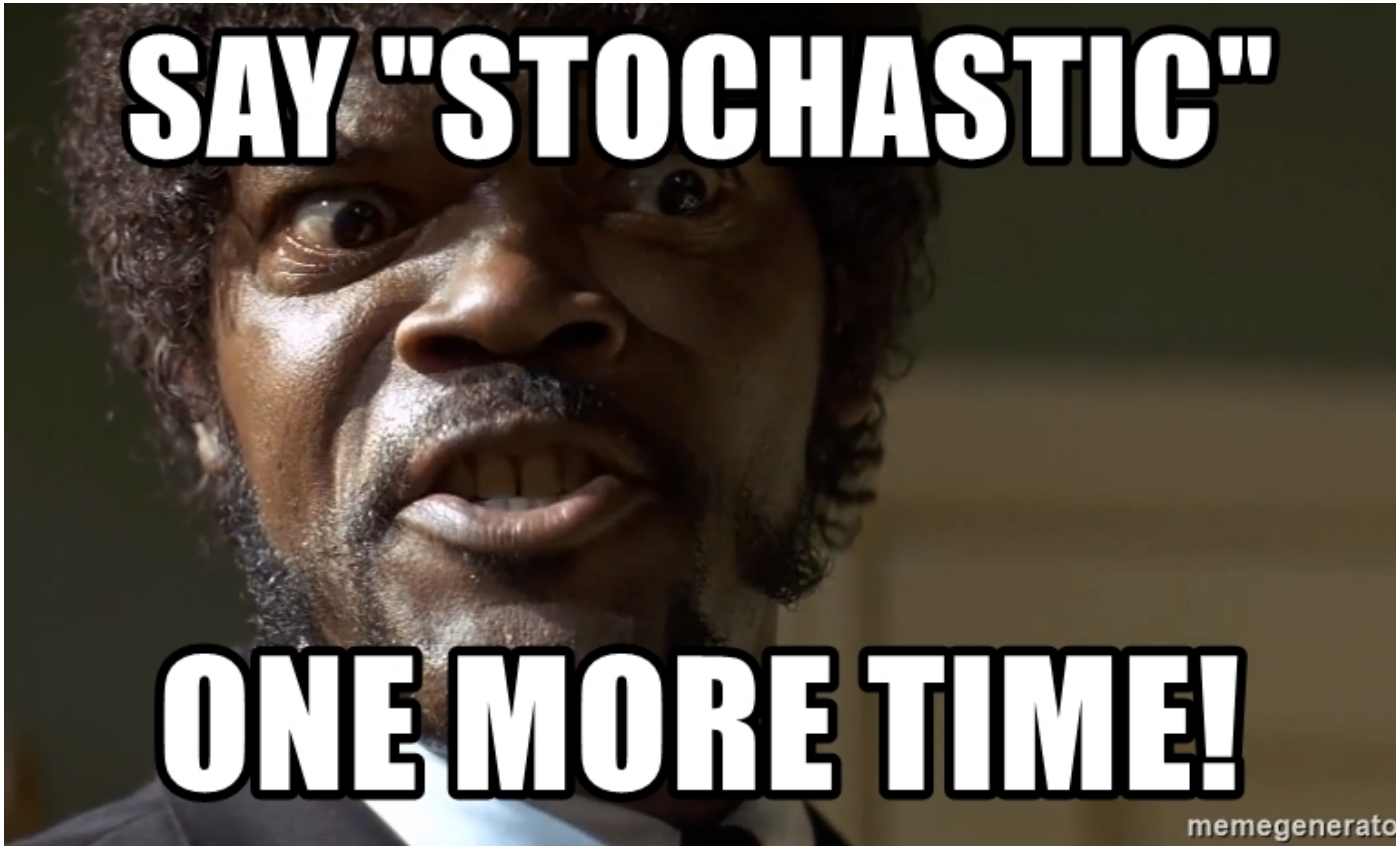
- Operated on singular data points

- Supervised Machine Learning means labeling

Google/OpenAI/etc.

Math is hard!

Source: https://www.instagram.com/p/BLN6MsWBf4W/ (sorry not sorry)

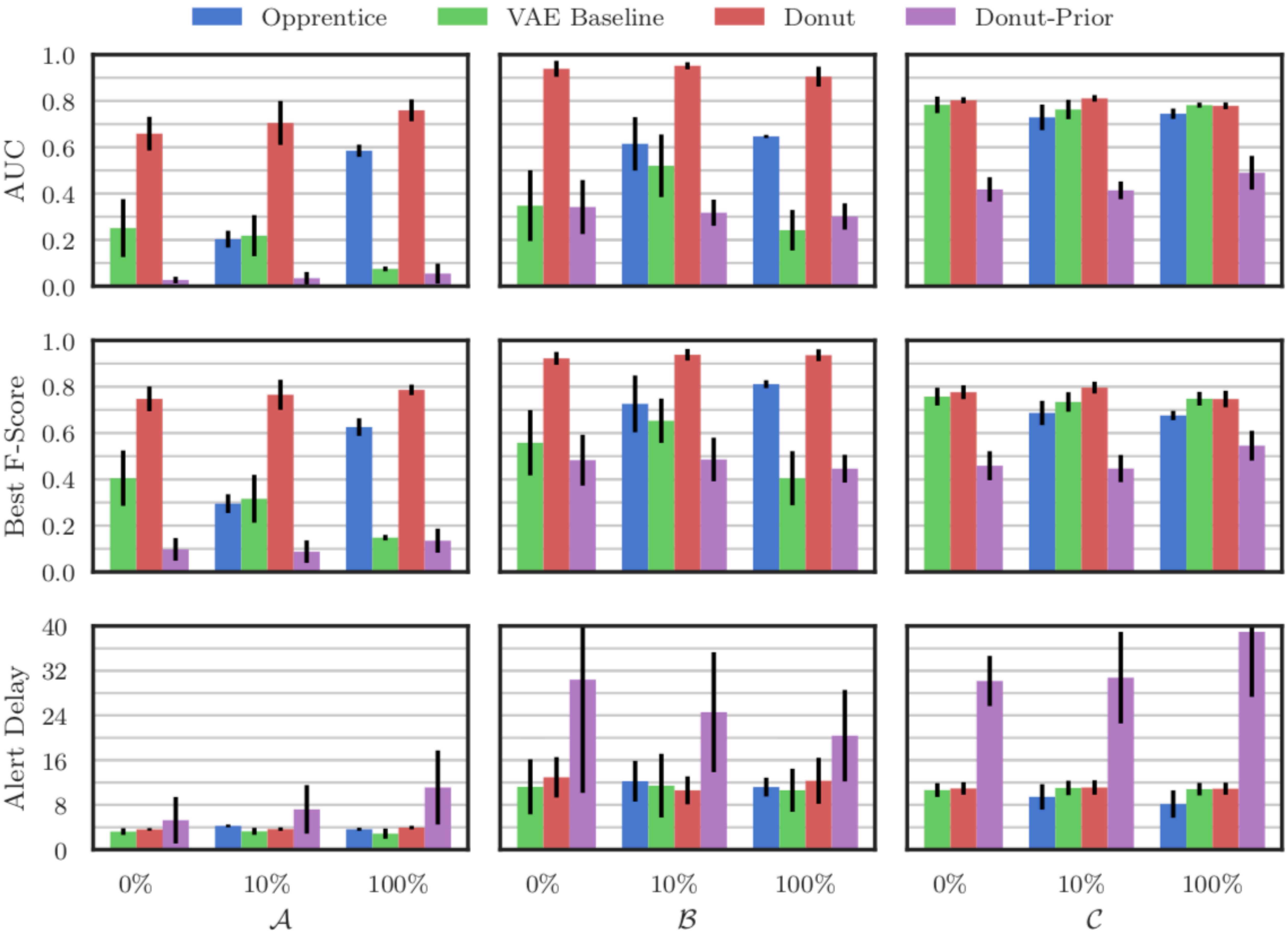


Source: memegenerator.net

Math is hard. Leave it to the experts.

- 🍩 Reducing algorithmic complexity of backpropagation from $O(K^3)$ to $O(K^2)$
- 🍩 Adam Optimizer
- 🍩 Reconstruction Probability

Results


Opprentice
VAE Baseline
Donut
Donut-Prior
AUC
Best F-Score
Alert Delay
1.0
0.8
0.6
0.4
0.2
0.0
40
32
24
16
8
0
0%
10%
100%
$\mathcal{A}$
$\mathcal{B}$
$\mathcal{C}$

“[...] it should be more important to have
an excellent F-score at a certain threshold than
to have just high but not
so excellent F-scores on most thresholds”

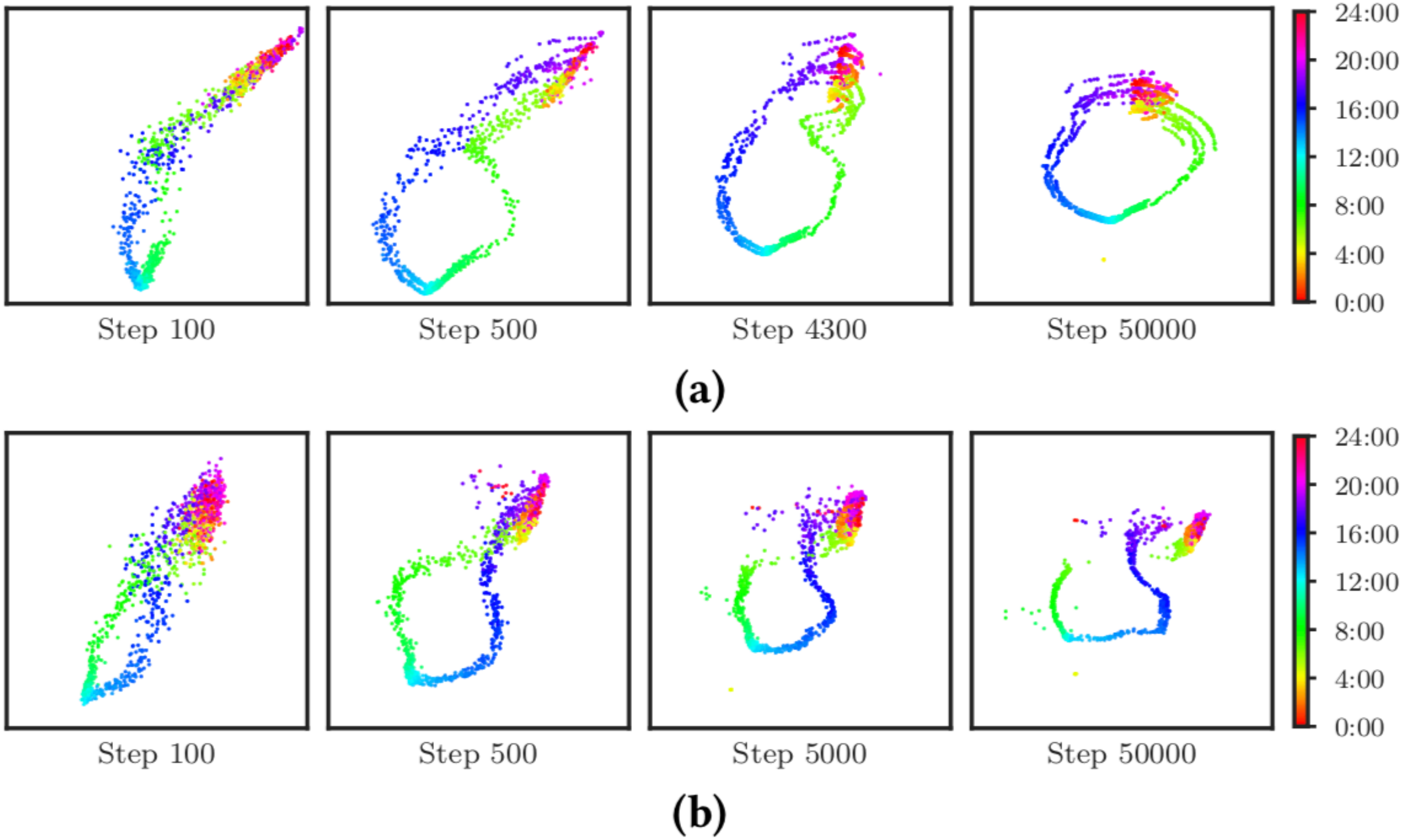
Closing Thoughts

Compare, Contrast, Reflect

Donut vs All

- 🍩 Does it generalize well?
- 🍩 Suboptimal convergence?
- 🍩 Does it scale well?

“For long-lasting anomalies, having the correct detection scores and raise alerts at first few minutes are sufficient in our context”



Suboptimal Convergence Happens

“[…] there is a convex optimization setting where Adam will not converge to the optimal solution, even if decreasing learning rates are used.”

–Sashank J. Reddi, Satyen Kale, Sanjiv Kumar

Please no questions

- EGADS: https://s.yimg.com/ge/labs/v2/uploads/kdd2015.pdf
- Opprentice: http://conferences2.sigcomm.org/imc/2015/papers/p211.pdf
- Reconstruction Probability: http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf

Fine. Ask your questions.

- EGADS: https://s.yimg.com/ge/labs/v2/uploads/kdd2015.pdf

- Opprentice: http://conferences2.sigcomm.org/imc/2015/papers/p211.pdf

- Reconstruction Probability: http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf