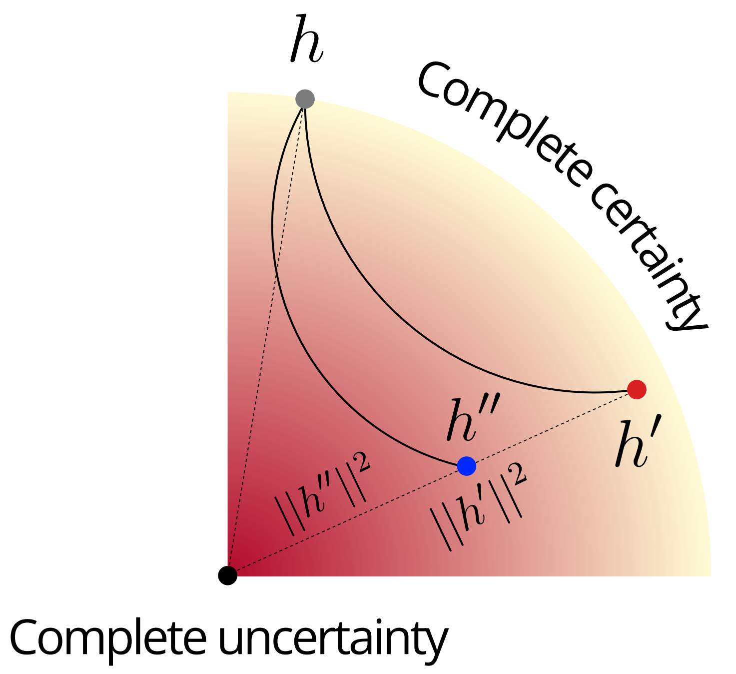


JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Are we certain it's anomalous?

Alessandro Flaborea[1], Bardh Prenkaj[1], Bharti Munjal[1,2]
Marco Aurelio Sterpa[1], Dario Aragona[1], Luca Podo[1],
Fabio Galasso[1]

1 What's the problem?

- Anomaly Detection in time series (ADTS) of events
- AD is ill-posed due to its unsupervised nature
- The literature neglects the trustfulness of the identified outliers = **uncertainty**
- End-to-end and data-driven uncertainty estimation is difficult due to the open-setness of anomalous events

2 How the literature approached ADTS

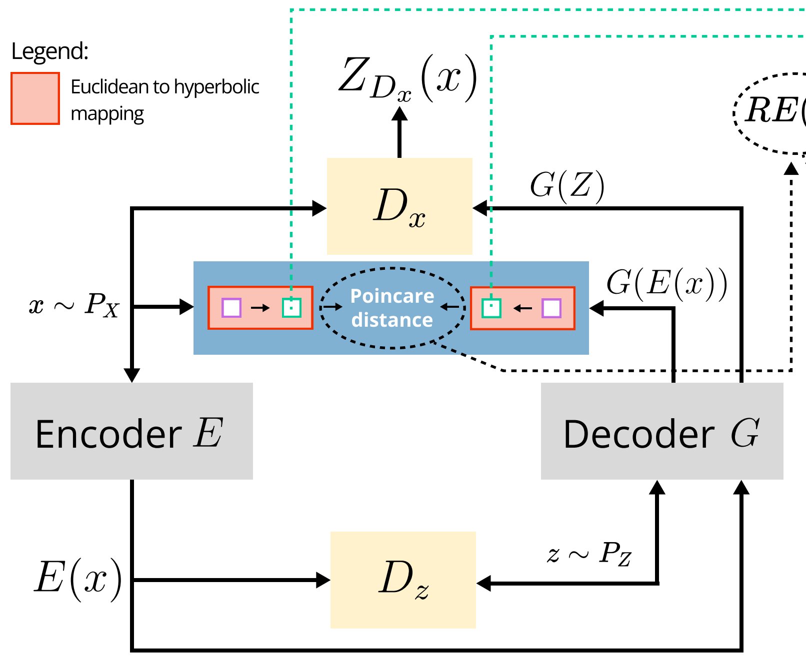
- Distance-based, density-based, prediction-based, and reconstruction-based detectors are baselines
- MAD-GAN [1] combines the discriminator output and reconstruction error
- BeatGAN [2] uses an encoder-decoder generator with a modified time-warping-based data augmentation
- TadGAN [3] uses a cycle-consistent GAN with a generator that computes reconstruction errors combined with the critic outputs

[1] Li *et al.* Mad-gan: Multivariate anomaly detection for time series data with generative adversarial net-works. In ICANN, 2019
[2] Zhou *et al.* Beatgan: Anomalous rhythm detection using adversarially generated time series. In IJCAI'19
[3] Geiger *et al.* Tadgan: Time series anomaly detection using generative adversarialnetworks. In BigData 2020

3 Proposed Approach

- HypAD is a reconstruction-based model that minimises the rec loss, given by a hyperbolic distance between the input and its reconstruction
- We use the Poincaré ball as the hyperbolic space

$$\exp_{\mathbf{0}}(x) = \tanh(||x||)\,\frac{x}{||x||}$$



$$RE(x) = \cosh^{-1}\left(1 + 2\,\frac{||\boxed{h} - \boxed{\tilde{h}}||^2}{(1 - ||h||^2)(1 - ||\tilde{h}||^2)}\right)$$

Grows **exponentially** as we move away from the origin of the Poincaré ball

- HypAD predicts either a matched reconstruction or an unmatched reconstruction towards the origin $U(x) = 1 - ||\tilde{h}||^2$
- Integrate $U(x)$ into the anomaly score

$$s_u(x) = Z_{RE}(x) \cdot Z_{D_x} \cdot (1 - U(x))$$

4 Take-away lessons

- Novel key idea of detectable anomaly

> *"A detectable anomaly is one where the model is certain, but it predicts wrongly"*

- HypAD outperforms SoA in univariate and multivariate time series

	AE	LstmAE	ConvAE	TadGAN	HypAD
MSL	0.199	0.317	0.300	0.500	**0.565**
SMAP	0.270	0.318	0.292	0.580	**0.643**
A1	0.283	0.310	0.301	**0.620**	0.610
A2	0.008	0.023	0.000	**0.865**	0.670
A3	0.100	0.097	0.103	**0.750**	0.670
A4	0.073	0.089	0.073	**0.576**	0.470
Art	0.283	0.261	0.289	0.420	**0.777**
AdEx	0.100	0.130	0.129	0.550	**0.663**
AWS	0.239	0.223	0.254	**0.670**	0.630
Traf	0.088	0.136	0.082	0.480	**0.570**
Tweets	0.296	0.299	0.301	0.590	**0.670**

Table 1: F1 scores on univariate datasets

	AE	LstmAE	ConvAE	TadGAN	HypAD
F	0.127	0.014	0.014	0.267	**0.333**
W	0.027	0.108	0.150	0.555	**0.610**
N	0.103	0.000	0.119	0.000	**0.333**
SW	0.000	0.049	0.048	**0.570**	0.364
MTC	0.049	0.035	0.035	0.222	**0.500**

Table 2: F1 scores on multivariate datasets

- Euclidean model (TadGAN) has a lot of FP; using hyperbolic + uncertainty (HypAD) recovers the detection of difficult anomalies