Are we certain it's anomalous?

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

[1] Sapienza University of Rome, Italy   [2] Technical University of Munich, Germany

{flaborea,prenkaj,aragona,podo,galasso}@di.uniroma1.it, munjalbharti@gmail.com

Abstract

The progress in modelling time series and, more generally, sequences of structured data has recently revamped research in anomaly detection. The task stands for identifying abnormal behaviors in financial series, IT systems, aerospace measurements, and the medical domain, where anomaly detection may aid in isolating cases of depression and attend the elderly. Anomaly detection in time series is a complex task since anomalies are rare due to highly non-linear temporal correlations and since the definition of anomalous is sometimes subjective.

Here we propose the novel use of Hyperbolic uncertainty for Anomaly Detection (HypAD). HypAD learns self-supervisedly to reconstruct the input signal. We adopt best practices from the state-of-the-art to encode the sequence by an LSTM, jointly learned with a decoder to reconstruct the signal, with the aid of GAN critics. Uncertainty is estimated end-to-end by means of a hyperbolic neural network. By using uncertainty, HypAD may assess whether it is certain about the input signal but it fails to reconstruct it because this is anomalous; or whether the reconstruction error does not necessarily imply anomaly, as the model is uncertain, e.g. a complex but regular input signal. The novel key idea is that a detectable anomaly *is one where the model is certain but it predicts wrongly. HypAD outperforms the current state-of-the-art for univariate anomaly detection on established benchmarks based on data from NASA, Yahoo, Numenta, Amazon, and Twitter. It also yields state-of-the-art performance on a multivariate dataset of anomaly activities in elderly home residences, and it outperforms the baseline on SWaT. Overall, HypAD yields the lowest false alarms at the best performance rate, thanks to successfully identifying detectable anomalies.*

1. Introduction

Anomaly detection stands for detecting outliers (anomalies) in data, i.e. points that deviate significantly from the distribution of the data. Outlier detection, however, is an
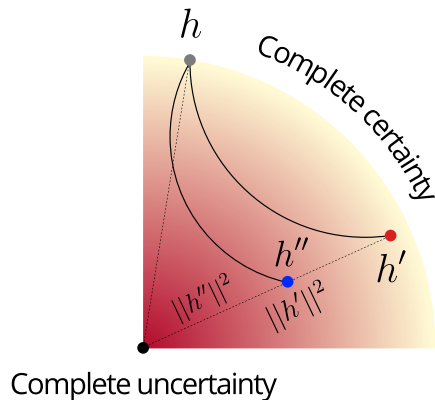

Figure 1. HypAD detects anomalies by the joint use of reconstruction error and uncertainty, learning both aspects end-to-end. In the illustration, the colored circular sector represents the hyperbolic Poincaré ball, where the radius distance of data embeddings is their degree of certainty (points on the circumference are most certain). h is the hyperbolic mapping of the input signal, which HypAD attempts to match by the reconstruction h'. Thanks to hyperbolic neural networks and their exponentially larger penalization for errors at high certainty, HypAD learns to prefer signal reconstructions such as h'', i.e. with the same amount of error as h' (the same angle and cosine distance) but smaller radius $\|h''\|^2$ and thus higher uncertainty. HypAD uses both the reconstruction error and the uncertainty to identify detectable anomalies, where the model is certain but the prediction is wrong, i.e. it knows what to expect but something anomalous occurs.

under-specified and consequently ill-posed task due to its inherent unsupervised nature. Anomaly detection strategies such as distance-based [16, 22], density-based [9, 55], and subspace-based methods [33, 40] have been pioneers in the literature. Additionally, autoencoders [11, 59], and adversarial networks [21] have given a substantial contribution. However, the literature neglects assessing the trustfulness of the predicted outcomes, namely their uncertainty.

Uncertainty is a measure of model confidence, which may be learnt from the data [34, 38] or by the use of extra instances [18]. Uncertainty estimation has been a long-standing challenge in machine learning. Most recently, it

has been successfully adopted to improve performance on object detection [32, 37, 45, 54], pose estimation [64], unsupervised and self-supervised learning [29, 58, 63]. Yet, end-to-end and data-driven uncertainty estimation for anomaly detection remains a complex task. This complexity is further exacerbated by the rarity of anomalous events which, generally, are not present during the learning (training) phase - i.e., open set problem - leading deep learning models to be overconfident when predicting them [27].

In this work, we propose a novel model based on Hyperbolic uncertainty for Anomaly Detection, which we dub HypAD. We leverage the current state-of-the-art technique for anomaly detection in univariate time series, TadGAN [21]. TadGAN detects anomalies by attempting to reconstruct the input signal, making use of an LSTM sequence encoding and two GAN critics, cf. Sec. 3.1. We introduce uncertainty into the anomaly detector: i.e., we map the input and the reconstructed signal into a hyperbolic space, where the signals additionally have an uncertainty score; and we train the novel embeddings end-to-end with a Poincaré distance loss, cf. Sec. 3.2.

HypAD uses uncertainty to discern whether the reconstruction error is large because the signal is anomalous, or simply because the model cannot reconstruct it well. In the former case, HypAD is certain about the reconstruction (e.g. most signal is well-behaved and the model expects known patterns) but its reconstruction is wrong, as a part of the signal is anomalous. In the latter, HypAD downgrades its anomaly score because it is not certain about signal reconstruction. This may be because of a complex pattern, which the model did not have enough capacity to learn. The larger uncertainty indicates that the larger reconstruction error may be due to an anomaly or to a model failure in the reconstruction. (See discussion in Sec. 3.2.2.)

Thanks to uncertainty, HypAD outperforms the state-of-the-art univariate anomaly detector TadGAN [21] on the established univariate benchmarks of NASA, Yahoo, Numenta Anomaly Benchmark [39], as well as on two multivariate datasets of daily activities in elderly home residences CASAS [12] and industrial water treatment plant SWaT [53]. As we show in experimental results in Sec. 4, reducing anomaly scores in uncertain cases also yields fewer false alarms (the model achieves the best F1 performance with larger precision).

The main contributions of this work are:

- We propose the first model for anomaly detection based on hyperbolic uncertainty;
- We propose the novel key idea of *detectable anomaly*: an instance is anomalous when the model is certain about it but wrong;
- We integrate the estimated uncertainty into a state-of-the-art univariate anomaly detector and outperform it on established univariate and multivariate datasets.

2. Related works

To the best of our knowledge, this is the first work to have combined anomaly detection with end-to-end and data-driven uncertainty estimation and the first work to have further proposed hyperbolic uncertainty for it. Previous work relates to ours from three main perspectives, which we review here: uncertainty estimation techniques, anomaly detection in time series, and hyperbolic neural networks.

2.1. Uncertainty Estimation Techniques

We identified two different strategies for approximating uncertainty. *Ensemble-based posterior approximation* uses several weak models to make naive predictions and combine them according to a consensus function into a more complex predictive model [14]. One of the most popular approaches to uncertainty estimation based on ensembles is Monte Carlo (MC) Dropout. It drops neurons on every layer during training and test phases [18].

Generative models for aleatory modelling use an additional latent variable to make stochastic predictions and evaluate the uncertainty of the model. Generative Adversarial Networks (GANs) [24] play a min-max game where the discriminator needs to distinguish between real examples and the generated outcomes. GANs have state-of-the-art performances and we build on top of that by attaching hyperspace mapping layers to estimate the uncertainty of the model. Another interesting approach to estimate uncertainty is by using energy-based models [15, 57, 67]. They learn an energy function that models the compatibility of the input and the output. Our method transcends energy-based models because the integrated hyperbolic uncertainty mechanism does not suffer from cold- nor warm-start problems [68] which undermine the training complexity [66].

Besides the above methods of approximating uncertainty, in the past, baselines like Hotelling's T^2 [28] exploited the Mahalanobis distance that incorporates uncertainty by measuring the (co-)variance, with the deviation (i.e., being anomalous or not). Nevertheless, current deep learning models in anomaly detection do not exploit uncertainty to calibrate their overconfident predictions for unseen data (anomalies) that lie outside the manifold of normal instances. Hence, an end-to-end and data-driven uncertainty estimation technique for anomaly detection is necessary.

2.2. Anomaly Detection in Time Series

We identified five categories of methods proposed in the literature for anomaly detection in time series. *Distance-based* outlier detectors consider the distance of a point from its k-nearest neighbours [4, 22, 36]. *Density-based* methods [9, 17, 26, 55, 61, 62, 65] take into account the density of the point and its neighbours. *Prediction-based methods* [2, 8, 31] calculate the difference between the predicted

value and the true value to detect anomalies. *Reconstruction based* methods [1, 3, 52] compare the input signal and the reconstructed one in the output layer typically by using autoencoders. These methods assume that anomalies are difficult to reconstruct and are lost when the signal gets mapped to lower dimensions. Thus, a higher reconstruction error means a higher anomaly score. [52] uses an LSTM autoencoder for multi-sensor anomaly detection. [11, 59] use an ensemble of autoencoders to boost performances by focusing on learning the inlier characteristics at each iteration. [46] uses a hierarchical variational auto-encoder with two stochastic latent variables to learn the temporal and inter-metric embeddings for multivariate data. SISVAE [44] uses a variational auto-encoder with a smoothness-inducing prior over possible estimations to capture latent temporal structures of time series without relying on the assumption of constant noise. Recently, *GANs* have been employed to detect anomalies in time series data. Our method also lies in this category. MAD-GAN [43] combines the discriminator output and reconstruction error to detect anomalies in multivariate time series. BeatGAN [69] uses an encoder-decoder generator with a modified time-warping-based data augmentation to detect anomalies in medical ECG inputs. TadGAN [21] uses a cycle-consistent GAN architecture with an encoder-decoder generator and proposes several ways to compute reconstruction error and its combination with the critic outputs. We build on top of TadGAN's architecture by incorporating the hyperbolic mapping layer to the reconstructed time-windows to assess the uncertainty of the detector.

2.3. Hyperbolic Neural Networks

Deep representation learning in hyperspaces has gained momentum after the pioneering work of hyperNNs [19] that generalizes Euclidean operations to their counterparts in hyperspace. The authors propose analog counterparts in the hyperspace of neural network components such as fully connected (FC) layers, multinomial logistic regression (MLR), and recurrent neural networks. Furthermore, methods like Einstain midpoint [25] and Fréchet mean [50] propose different ways of aggregating features in hyperspace. The work in [60] extends hyperNN and proposes Poincaré split/concatenation operations, generalizing the convolutional layer to hyperspace. [10, 47] propose hyperbolic graph neural networks, leveraging hyperNNs.

Thus formulated, hyperNNs have mainly been adopted to improve performance by leveraging hierarchies and uncertainty in zero-shot learning [48], re-identification [35], and action recognition [49]. Of particular interest, [63] has leveraged hyperNNs to model a hierarchy of actions from unlabeled videos. To the best of our knowledge, this is the first work to have applied hyperNNs for sequence modelling with the goal of anomaly detection.
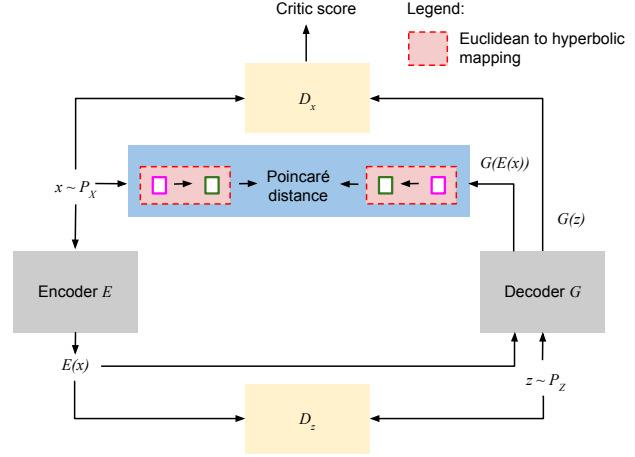

Figure 2. Overall architecture of our proposed model HypAD. The model integrates hyperNNs [19] with an LSTM based encoder-decoder trained with two *GAN*-based discriminators, D_x and D_z. HypAD maps the input signal x and the output of the decoder $(G(E(x)))$ to the Poincaré ball model of hyperbolic spaces, shown as dotted red edge box with red background, to get the corresponding hyperbolic embeddings. The two embeddings are then compared using Poincaré distance, as described in Sec. 3.2.1. Solid magenta and green boxes denote the input and output of the hyperbolic mapping, respectively.

3. Method

In this section, we first discuss best practises in anomaly detection (Sec. 3.1); then we detail the proposed hyperbolic uncertainty and its use for detecting anomalies (Sec. 3.2); finally, we discuss the motivation for it (Sec. 3.3).

3.1. Background

The current state-of-the-art in univariate anomaly detection is a reconstruction-based technique [21] which additionally leverages a GAN critic score. TadGAN encodes the input data x to a latent space and then decodes the encoded data. This encoding-decoding operation requires two mapping functions $E : X \to Z$ and $G : Z \to X$. The reconstruction operation can be given as $x \to E(x) \to G(E(x)) \to \widetilde{x} \approx x$. TadGAN leverages adversarial learning to train the two mappings by using two adversarial critics D_x and D_z. The goal of D_x is to distinguish between the real and the generated time series, while D_z measures the performance of the mapping into latent space. The model is trained using the combination of Wasserstein loss [5] and Cycle consistency loss [70]. TadGAN computes the reconstruction error $RE(x)$ between x and $\widetilde{x}$ using three types of reconstruction functions: **i.** Point-wise difference: considers the difference of values at every timestamp; **ii.** Area difference: is applied on signals of fixed lengths and measures the similarity between local regions; **iii.** Dynamic time warping: additionally handles time gaps between the two
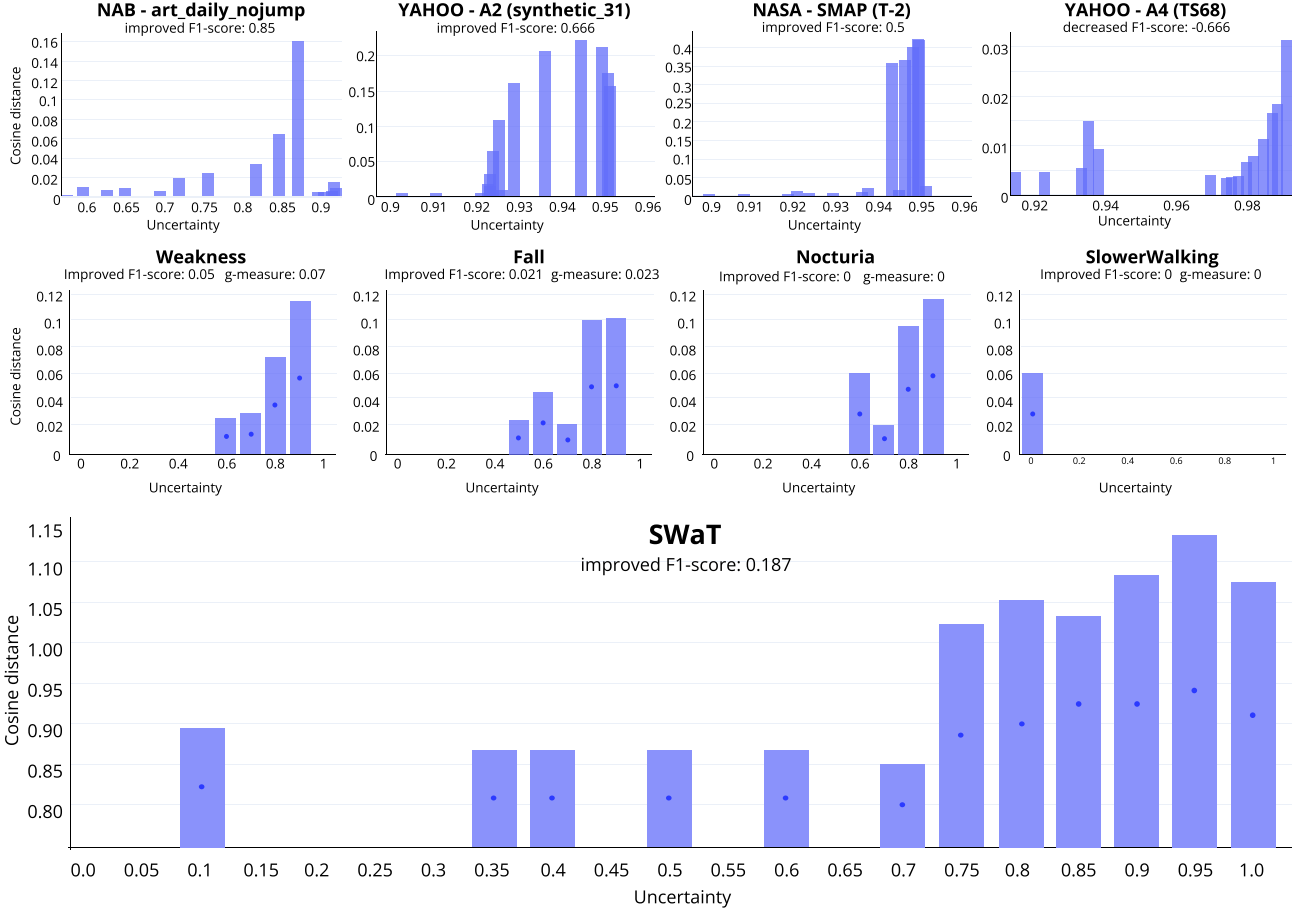
Figure 3. Bar plot of the average cosine distance of all the datasets for specific intervals of uncertainty (see Sec. 3.3). The first two rows contain three plots corresponding to the signals that report the best improvement in terms of F1-score (g-measure) and one plot (the last column) that corresponds to the signal with the worst improvement. Notice that even the signals with the worst improvement follow the same increasing trend. Because SWaT contains a single long-term signal, we report only its corresponding bar-plot.

signals to calculate the reconstruction error. To calculate the anomaly score, TadGAN first normalises the reconstruction error and critic scores by subtracting the mean and dividing by standard deviation. The normalised scores, $Z_{RE}(x)$ and $Z_{D_x}(x)$, are then combined using their product:

$$s_p(x) = Z_{RE}(x) \cdot Z_{D_x}(x) \tag{1}$$

3.2. Hyperbolic uncertainty for Anomaly detection (HypAD)

We propose a novel model for anomaly detection in time series based on hyperbolic uncertainty. HypAD is a reconstruction-based model and minimises the reconstruction loss, given by a measure of the hyperbolic distance between the input signal and its reconstruction. In hyperbolic space, errors are exponentially larger when predictions are certain. Therefore, HypAD tends to predict either certain correct reconstructions or uncertain possibly mistaken reconstructions. This leads, as we discuss in Sec. 3.3, to a novel definition of detectable anomaly: i.e. the case of a

large reconstruction error with high certainty.

3.2.1 Hyperbolic Reconstruction Error

The proposed HypAD, illustrated in Fig. 2, integrates hyperbolic neural networks into the reconstruction-based architecture of TadGAN. In HypAD the input signal x is first passed through an encoder, then followed by a decoder subnetwork. The output of the decoder $G(E(x))$ as well as the original signal x are mapped to the hyperspace, shown as the red edge box with red background.

An n-dimensional hyperspace is a Riemannian geometry with a constant negative sectional curvature. As in [35, 63], we adopt the Poincaré ball model of hyperspaces, given by the manifold $\mathbb{D}_n = \{x \in R^n : \|x\| < 1\}$ endowed with the Riemannian metric $g^{\mathbb{D}}(x) = \lambda_x^2 g^E$, where $\lambda_x = \frac{2}{1-\|x\|^2}$ is the *conformal factor* and $g^E = I_n$ is the Euclidean metric tensor. For details, see [41, 42].

To map x and $G(E(x))$ to the Poincaré ball, we leverage an *exponential map* centered at $\mathbf{0}$ [63]. This is followed by

	Univariate											Multivariate					
	NASA		Yahoo				NAB					CASAS					SWaT
	SMAP	MSL	A1	A2	A3	A4	Art	AdEx	AWS	Traf	Tweets	F	MTC	W	SW	N	
Num. of signals	53	27	67	100	100	100	6	5	17	7	10	1	1	1	1	1	1
Num. of anomalies	67	36	178	200	939	835	6	11	30	14	33	2	2	4	2	2	33
Point anomalies	0	0	68	33	935	833	0	0	0	0	0	0	0	0	0	0	0
Collective anomalies	67	36	110	167	4	2	6	11	30	14	33	2	2	4	2	2	33
Num. of anomaly points	54,696	7,766	1,699	466	943	837	2,418	795	6,312	1,560	15,651	99	239	1,248	276	1,060	10786
Percentage of total	9.7%	5.8%	1.8%	0.3%	0.6%	0.5%	10%	9.9%	9.3%	9.8%	9.8%	0.7%	1.7%	8%	2.4%	6.3%	10.05%
Num. out of distribution	18,126	642	861	153	21	49	123	15	210	86	520	-	-	-	-	-	-
Num. of instances	$\sim 563k$	$\sim 132k$	$\sim 95k$	$\sim 142k$	168k	168k	$\sim 24k$	$\sim 8k$	$\sim 68k$	$\sim 16k$	$\sim 159k$	$\sim 54k$	$\sim 276k$	$\sim 39k$	$\sim 39k$	$\sim 39k$	$\sim 189k$
Synthetic?	No	No	No	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No	No	No	No

Table 1. Overview of selected univariate and multivariate datasets and their characteristics, grouped by the sources of signals (NASA, Yahoo, NAB, CASAS and SWaT). See Sec. 4.1 for details.

a hyperbolic feed-forward layer [19] to estimate the corresponding hyperbolic embeddings h and $\widetilde{h}$, solid green boxes in Fig. 2. Finally, the two hyperbolic embeddings are compared using the Poincaré distance, formulated as follows:

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

where $\|h\|^2$ and $\|\widetilde{h}\|^2$ are the distances of the embeddings from the center of the Poincaré ball. The same reconstruction error function $RE(x)$, is used at train and inference.

3.2.2 Hyperbolic Uncertainty

A key property of the Poincaré ball is that the distance between two points grows exponentially as we move away from the origin. This means that an erroneous reconstruction towards the circumference of the disk is penalised exponentially more than an erroneous reconstruction close to the centre. This leads to the useful tendency of HypAD to either predict a matched reconstruction ($\widetilde{h}$ and h are close by) or an unmatched reconstruction towards the origin ($\widetilde{h}$ and h are far away, $\|h\|^2$ and $\|\widetilde{h}\|^2$ are small), in order to minimize Eq. (2). Hence, the distance of the reconstruction to the origin provides a natural estimate of the model's uncertainty, referred to as hyperbolic uncertainty, $U(x)$, thus formulated:

$$U(x) = 1 - \|\tilde{h}\|^2 \quad (3)$$

The smaller the distance from the origin, the more uncertain the model.

3.2.3 Combining Hyperbolic Uncertainty with Reconstruction Error and Critic Score

Hyperbolic uncertainty $U(x)$ is integrated into the anomaly score as follows:

$$s_u(x) = Z_{RE}(x) \cdot Z_{D_x}(x) \cdot (1 - U(x)) \quad (4)$$

Eq. (4) brings together the reconstruction error $Z_{RE}(x)$ (the larger the error, the more likely the anomaly) with the critic

score $Z_{D_x}(x)$ (larger critic scores point to anomalies) and the model certainty: $1 - U(x)$. The simple multiplication formulation of the certainty of the model reduces the scores of anomalies when HypAD is not confident of the reconstructions. While being simple, this outperforms the current state-of-the-art, as we show in Sec. 4.

3.3. Motivation for HypAD

HypAD takes motivation from a key idea: *a detectable anomaly is one where the model is certain, but it predicts wrongly*. In other words, if the model encounters a known pattern, which it knows how to reconstruct, then it will call it anomalous if the reconstruction does not match the input signal. The principled formulation of hyperbolic uncertainty is critical towards this goal: HypAD predicts a reconstruction as uncertain if it is doubtful that it may be wrong.

Fig. 3 illustrates this key concept for all the datasets. The first, second and third rows correspond to the univariate, U-CASAS and SWaT datasets respectively. The bar plot depicts the average cosine distances between the input signals and their reconstructions, against specific intervals of uncertainty, along the x-axis. The higher the cosine distances, the more distinct the reconstruction is from the provided signals[1]. Note that for the first two rows, the initial three plots (columns) correspond to the signals that report the best improvement in F1-score and the last plot corresponds to the signal with the worst improvement. A single bar plot is reported for SWaT because this consists of a single long-term signal. Observe, in the second row, how HypAD learns to correctly assign higher uncertainty to wronger estimates for the cases of Fall, Weakness, and Nocturia. For the last signal SlowerWalking, HypAD fails to learn a meaningful uncertainty and labels all reconstructions as certain. It is however notable that the representation is still interpretable and the case of failure discernible. Trends are similar for the cases of Univariate and SWaT datasets.

---

[1]The Poincaré ball model is conformal to the Euclidean space and it preserves the same angles [60].

4. Results

We compare HypAD against the best univariate anomaly detector TadGAN [21] on a benchmark of 11-time series and extend the comparison to 2 multivariate sensor datasets: one comprising a water treatment plant and one comprising daily activities in elderly residences. First, we introduce the benchmarks (Sec. 4.1), then we compare against baselines and the state-of-the-art (Sec. 4.2), finally, we conduct ablative studies on the importance of uncertainty for performance and the reduction of false alarms (Sec. 4.3).

4.1. Datasets, metrics and experimental setup

Tab. 1 summarizes the main characteristics of the datasets, which we coarsely divide into univariate, main test bed of our baseline TadGAN [21], and multivariate, to which we extend the comparison. In the table, we report the sources of signals (NASA, Yahoo, NAB, CASAS and SWaT) and the datasets within each source (SMAP, MSL, A1, etc.), cf. detailed description later in the section.

In the table, the *number of signals* is the number of time series within the datasets. Note that univariate datasets are composed of multiple signals, while multivariate only comprise single large time series. The *number of anomalies* counts the instances, within the time series, labeled as anomalous. These are further detailed as *point anomalies*, anomalous values at a specific point in time, or *collective anomalies*, sets of contiguous times that are altogether anomalous. Yahoo is the sole dataset with point anomalies and with *synthetic* sequences A2, A3, and A4. We also report the *percentage of total* anomalous points and, following [21], for the univariate datasets, the *number of out-of-distribution* points, exceeding the means by more than 4σ.

Univariate datasets. *NASA* includes two spacecraft telemetry datasets based on the Mars Science Laboratory (MSL) and the Soil Moisture Active Passive (SMAP) signals. The former consists of scientific and housekeeping engineering data taken from the Rover Environmental Monitoring Station aboard the Mars Science Laboratory. The latter includes measurements of soil moisture and freeze/thaw state from space for all non-liquid water surfaces globally within the top layer of the Earth. The *Yahoo* datasets are based on real production traffic to Yahoo computing systems. Additionally, we consider three synthetic datasets coming from the same source. The dataset tests the detection accuracy of various anomaly types including outliers and change points. The synthetic dataset consists of time-series with varying trends, noise and seasonality. The real dataset consists of time-series representing the metrics of various Yahoo services. *Numenta Anomaly Benchmark (NAB)* is a well-established collection of univariate time-series from real-world application domains. To be consistent with [21], we analyse Art, AdEx, AWS, Traf, and Tweets from the original collection.

Multivariate datasets. We consider for analysis the SWaT [23, 53] and CASAS [12, 13] datasets. *SWaT* is collected from a cyber-physical system testbed that is a scaled-down replica of an industrial water treatment plant. The data was collected every second for a total of 11 days, where for the first few days the system was operated normally, while for the remaining days, certain cyber-physical attacks were launched. Following [17], we sample sensor data every 5 seconds. *CASAS* is a collection of two weeks of sensor data from retirement homes. Each sensor reading has a label attached to it, according to the activity of elderly people recognised by human annotators. The 5-time series are a collection of activities, organised by the pre-established medical conditions: Falling, More Time in Chair, Weakness, Slower Walking, and Nocturia for nightly time visits. Although sensor readings give fine-grained information, we are interested in creating daily patient profiles. Hence, we collapse them into a single engendered activity with a start and end time for each consecutive sensor signal. We then create a time matrix structure $\mathcal{M}^{|d|,1440}$ where $|d|$ is the number of days the patient is monitored and 1440 represent the total number of minutes in a day. $\mathcal{M}[i,j]$ represents the label performed in minute j of day i. Lastly, we densify each value $\mathcal{M}[i,j]$ with contextual and duration information corresponding to the label therein.

Finally, we create train and test splits of this data, safeguarding the sequentiality of observations, and gathering the few anomalies into the test set for evaluation only. We name this Unsupervised-CASAS, dubbed U-CASAS, differing from the original CASAS [12, 13] since the latter interleaves anomalous sensor readings with normal instances, thus breaking the sequentiality of anomalies. For each anomalous day encountered in the test set, we pad two days prior to it and two after. If the padding overlaps with another anomalous day, then we concatenate[2] them and perform the padding procedure again. We delete the days assigned to the test set from the overall dataset and assign the rest as training. Moreover, since we employ a time-related strategy, we create time windows of 30 actions to detect anomalies.

Metrics. Being all the enlisted datasets highly unbalanced, accuracy is misleading. Therefore, as done in [63], we use the F1 score to account for this challenge. Notice that we do not use the cumulative F1 score as proposed in [20] to evaluate the performances because not all datasets contain anomalous events; rather they contain anomalous data points. Based on [30], for the univariate and the CASAS datasets, we penalize high false positive rates and encourage the detection of true positives in a timely fashion. Since anomalies are rare events and come in collective sequences in real-world applications, we proceed as follows:

---

[2]The concatenation procedure merges common sequences. If the sequence contains more than one anomalous day within the two-day padding window, the padded sequence gets collapsed into a single one.

| | NASA | | YAHOO | | | | NAB | | | | | |
	MSL	SMAP	A1	A2	A3	A4	Art	AdEx	AWS	Traf	Tweets	F1 ($\mu \pm \sigma$)
AE	0.199	0.270	0.283	0.008	0.100	0.073	0.283	0.100	0.239	0.088	0.296	0.176 ± 0.099
LstmAE	0.317	0.318	0.310	0.023	0.097	0.089	0.261	0.130	0.223	0.136	0.299	0.200 ± 0.103
ConvAE	0.300	0.292	0.301	0.000	0.103	0.073	0.289	0.129	0.254	0.082	0.301	0.212 ± 0.096
TadGAN	0.500	0.580	**0.620**	**0.865**	**0.750**	**0.576**	0.420	0.550	**0.670**	0.480	0.590	0.600 ± 0.115
HypAD (proposed)	**0.565**	**0.643**	0.610	0.670	0.670	0.470	**0.777**	**0.663**	0.630	**0.570**	**0.670**	**0.631 ± 0.075**

Table 2. Results on the univariate datasets from NASA, YAHOO and NAB, measured in terms of F1-score.

| | Fall | | Weakness | | Nocturia | | Slower-walking | | More time in chair | | G ($\mu \pm \sigma$) | F1 ($\mu \pm \sigma$) |
	G	F1	G	F1	G	F1	G	F1	G	F1		
LstmAE	0.085	0.014	0.182	0.108	0.000	0.000	0.158	0.049	0.133	0.035	0.112 ± 0.064	0.041 ± 0.037
AE	0.139	0.127	0.033	0.027	0.116	0.103	0.000	0.000	0.158	0.049	0.089 ± 0.062	0.061 ± 0.047
ConvAE	0.086	0.014	0.284	0.150	0.251	0.119	0.158	0.048	0.134	0.035	0.183 ± 0.074	0.073 ± 0.052
TadGAN	0.222	0.267	0.570	0.555	0.000	0.000	**0.630**	**0.570**	0.267	0.222	0.338 ± 0.233	0.323 ± 0.216
HypAD (proposed)	**0.447**	**0.333**	**0.660**	**0.610**	**0.447**	**0.333**	0.470	0.364	**0.577**	**0.500**	**0.520 ± 0.095**	**0.428 ± 0.123**

Table 3. Results for the U-CASAS multivariate datasets, measured in terms of F1-score and G-measure.

1. We record a true positive (TP) if any predicted window overlaps a true anomalous window.
2. We record a false negative (FN) if a true anomalous window does not overlap any predicted window.
3. We record a false positive (FP) if a predicted window does not overlap any true anomalous region.

For U-CASAS, we also measure the g-measure which is the geometric mean of the product of recall (R) and precision (P), being a robust metric when classes are imbalanced [13].
Baselines. We include the following strategies as our baselines in this paper:

- *AE* [7] - We use a six-layer dense autoencoder.
- *ConvAE* [51] - We have three layers of convolutional encoding interleaved with max pooling. The decoder has a specular composition as the encoder where the de-convolution is aided by two-dimensional up-sampling layers.
- *LstmAE* [56] - We use a deep-stacked LSTM autoencoder with four layers. The first LSTM hidden and output vectors get passed to the second LSTM layer. The latent representation of the encoder gets then reconstructed in reverse order from the decoder.
- *TadGAN* [21] - We use a one-layer bidirectional LSTM for the generator E, and a two-layer bidirectional LSTM for G. For the critic D_z we use a fully connected layer, and two dense layers for D_x.

Implementation details. For the first three baselines, we set the number of epochs to 30, the batch size to 32, and the learning rate to 10^{-3}. For TadGAN[3], we set the epochs to 30, the batch size to 64, the window width to 100, the learning rate to 5×10^{-4}, and the iteration for the critic to 5. We use Adam as the optimizer to train all the baselines. For our proposed method HypAD, we took inspiration from the PyTorch implementation of TadGAN. We leave the architecture of TadGAN unvaried, and incorporate the hyper-

---

[3]Pytorch implementation available at https://github.com/arunppsg/TadGAN

bolic transformation as in [63]. The hyperparameters are the same as in the original paper, using Riemannian Adam as optimizer.

4.2. Comparison to the state of the art

HypAD defines a new state-of-the-art performance for univariate anomaly detection by having the highest average F1-scores of 0.631. In Tab. 2, HypAD outperforms the current best technique, TadGAN, by 5.17%, as well as all baselines by a large margin. In the Table, the column *F1* reports mean μ and standard deviation σ over all datasets. Looking at $\sigma = 0.075$, HypAD also appears as the most consistent performer. Considering the F1-score, the largest performance gain of HypAD Vs TadGAN are on NAB and NASA datasets, while it is outperformed more largely on the A2, A3 and A4 Yahoo datasets, the synthetic ones.

In Tab. 3, we extend the evaluation of HypAD to the multivariate U-CASAS dataset. We cannot include Isudra [13] because the underlying architecture uses a small amount of labels to select parameters for the execution. Additionally, Isudra is trained in a supervised fashion differing from all the other methods reported here. For completeness, we report the g-measure [13] and its average across all datasets. As shown in the table, HypAD surpasses TadGAN by 32.51% in terms of the average F1-score.

Finally, in Tab. 4, we extend the comparison to the multivariate SWaT dataset. Here, following [17], we report the precision and recall in addition to the F1-score. HypAD outperforms the baseline TadGAN (0.753 Vs. 0.722), but both techniques are behind the current SoA on multivariate anomaly detection, NSIBF [17]. HypAD achieves its highest F1 performance at the highest precision (0.996) among all other methods. This confirms that HypAD detects anomalies which it is certain about, i.e. when it *understands* the time series and it is certain about what to expect, but cannot reconstruct the input signal due to an anomaly.

Model	Precision	Recall	F1
EncDec-AD [52] [ICML-WorkShop16]	0.945	0.620	0.748
DAGMM [71][ICLR18]	0.946	0.747	0.835
OmniAnomaly [61] [KDD19]	0.979	0.757	0.854
USAD [6] [KDD20]	0.987	0.740	0.846
TadGAN [BigData20]	0.937	0.587	0.722
NSIBF [17] [KDD21]	0.982	**0.863**	**0.919**
HypAD (Ours)	**0.996**	0.605	0.753

Table 4. Results on the SWAT dataset, in terms of F1-score, including the corresponding precision and recall.
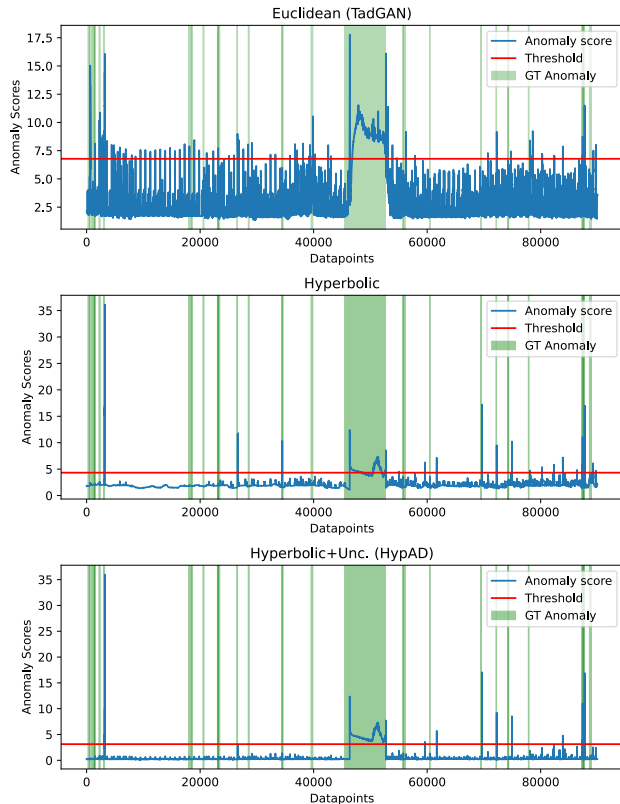

Figure 4. Qualitative Ablation on SWaT: Anomaly scores (blue points) that lie above the anomaly detection threshold (red line) but do not coincide with the ground-truth (green points) are FP. Euclidean model in plot 1 has many FP. The corresponding hyperbolic model in plot 2 is more precise but loses some TP, especially in the middle region. Integration of uncertainty helps to recover these points, increasing the TP and overall F-1 score (Sec 4.3).

4.3. Ablation Studies

In Tab. 5, we analyze the importance of the hyperbolic embedding and the use of uncertainty for anomaly detection on the univariate datasets, as well as the multivariate U-CASAS and SWaT. The first row shows the performance of the model in Euclidean space (average F1-score across datasets). This corresponds to TadGAN in Tab. 2, 3 and 4. In the second row, we report the performance for TadGAN with the inclusion of uncertainty via 100 Monte

	Univariate	U-CASAS	SWaT
Euclidean (TadGAN)	0.600	0.323	0.722
Euclidean (TadGAN) + MC dropout	0.618	0.339	0.734
Hyperbolic	0.604	0.397	0.566
Hyperbolic + Uncertainty (HypAD)	**0.631**	**0.428**	**0.753**

Table 5. Ablative evaluation on the importance of the hyperbolic mapping and the integration of uncertainty. Average F1-scores are reported for Univariate, U-CASAS and SWaT datasets.

Carlo dropouts. In the third row, we report performance for the hyperbolic TadGAN without including uncertainty (cf. Sec. 3.2.1). This improves marginally over the univariate (0.604 Vs. 0.600) and, more importantly, over the U-CASAS datasets (0.397 Vs. 0.323), but it decreases performance in SWaT (0.566 Vs. 0.722), which we further analyze in the following subsection. The complete proposed HypAD model, in the fourth row, improves consistently on both ablative variants. HypAD yields a 4.5% gain on the univariate datasets, 7.8% on U-CASAS, and 33% on SWaT. We infer that hyperbolic mapping and its associated uncertainty are fundamental for ameliorating anomaly detection over the Euclidean counterparts.

Qualitative Ablation on SWaT. In Fig. 4, we demonstrate qualitative ablation on the SWaT dataset, which consists of a single long signal. The three plots correspond to the three ablative variants of Tab. 5. In all plots, the blue and green points represent the predicted anomaly scores and the ground-truth anomalies, respectively. The red line denotes the anomaly detection threshold i.e. blue points above the red line are the predicted anomalies. FP are blue points above the red line threshold outside the green (ground-truth) anomalous regions. As shown in the figure, the Euclidean model (first plot) yields many FP; the hyperbolic model w/o uncertainty (second plot) reduces the number of FP substantially, but it also misses anomalies, esp. in the middle signal part. This explains the drop in the F-1 score for the SWaT dataset (cf. Table 5). Integrating hyperbolic uncertainty (proposed HypAD, third plot) recovers the detection of these anomalies. It increases TP but maintains the number of FP low, thus yielding the best F1 score of 0.753.

5. Conclusions

We have proposed a novel model for anomaly detection based on hyperbolic uncertainty, HypAD. The proposed hyperbolic uncertainty allows HypAD to self-adjust its output, encouraging the model to either predict a correct reconstruction or a less certain wrong one. This benefits anomaly detection in two ways: it provides better reconstructions of the signal (the deviation from those being anomalous) and it yields a measure of certainty. This is a novel viewpoint on anomaly detection: detectable anomaly instances are those which are certain but wrongly predicted.

References

[1] Ahmed Abdulaal, Zhuanghua Liu, and Tomer Lancewicki. Practical approach to asynchronous multivariate time series anomaly detection and localization. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, page 2485–2494. ACM, 2021. 3

[2] Subutai Ahmad, Alexander Lavin, Scott Purdy, and Zuha Agha. Unsupervised real-time anomaly detection for streaming data. *Neurocomputing*, 262, 06 2017. 2

[3] Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection usin reconstruction probability. In *Special Lecture on IE*, volume 2, 2015. 3

[4] Fabrizio Angiulli and Clara Pizzuti. Fast outlier detection in high dimensional spaces. In *European Conference on Principles of Data Mining and Knowledge Discovery*, pages 15–27. Springer, 2002. 2

[5] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 214–223. PMLR, 06–11 Aug 2017. 3

[6] Julien Audibert, Pietro Michiardi, Frédéric Guyard, Sébastien Marti, and Maria A. Zuluaga. Usad: Unsupervised anomaly detection on multivariate time series. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 3395–3404. ACM, 2020. 8

[7] Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML workshop on unsupervised and transfer learning*, pages 37–49. JMLR Workshop and Conference Proceedings, 2012. 7

[8] Seif-Eddine Benkabou, Khalid Benabdeslem, Vivien Kraus, Kilian Bourhis, and Bruno Canitia. Local anomaly detection for multivariate time series by temporal dependency based on poisson model. *IEEE Transactions on Neural Networks and Learning Systems*, 2021. 2

[9] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. Lof: identifying density-based local outliers. In *ACM sigmod record*, volume 29, pages 93–104. ACM, 2000. 1, 2

[10] Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. *Advances in neural information processing systems*, 32:4868–4879, 2019. 3

[11] Jinghui Chen, Saket Sathe, Charu Aggarwal, and Deepak Turaga. Outlier detection with autoencoder ensembles. In *Proceedings of the 2017 SIAM International Conference on Data Mining*, pages 90–98. SIAM, 2017. 1, 3

[12] Diane J Cook, Aaron S Crandall, Brian L Thomas, and Narayanan C Krishnan. Casas: A smart home in a box. *Computer*, 46(7):62–69, 2012. 2, 6

[13] Jessamyn Dahmen and Diane J Cook. Indirectly supervised anomaly detection of clinically meaningful health events from smart home data. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 12(2):1–18, 2021. 6, 7

[14] Thomas G Dietterich. Ensemble methods in machine learning. In *International workshop on multiple classifier systems*, pages 1–15. Springer, 2000. 2

[15] Yilun Du and Igor Mordatch. Implicit generation and modeling with energy based models. *Advances in Neural Information Processing Systems*, 32:3608–3618, 2019. 2

[16] Hongqin Fan, Osmar R Zaïane, Andrew Foss, and Junfeng Wu. A nonparametric outlier detection for effectively discovering top-n outliers from engineering data. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 557–566. Springer, 2006. 1

[17] Cheng Feng and Pengwei Tian. Time series anomaly detection for cyber-physical systems via neural system identification and bayesian filtering. In *The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 2858–2867. ACM, 2021. 2, 6, 7, 8

[18] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pages 1050–1059. PMLR, 2016. 1, 2

[19] Octavian Ganea, Gary Becigneul, and Thomas Hofmann. Hyperbolic neural networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. 3, 5

[20] Astha Garg, Wenyu Zhang, Jules Samaran, Ramasamy Savitha, and Chuan-Sheng Foo. An evaluation of anomaly detection and diagnosis in multivariate time series. *IEEE Transactions on Neural Networks and Learning Systems*, 33(6):2508–2517, 2021. 6

[21] Alexander Geiger, Dongyu Liu, Sarah Alnegheimish, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Tadgan: Time series anomaly detection using generative adversarial networks. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 33–43. IEEE, 2020. 1, 2, 3, 6, 7

[22] Amol Ghoting, Srinivasan Parthasarathy, and Matthew Eric Otey. Fast mining of distance-based outliers in high-dimensional datasets. *Data Mining and Knowledge Discovery*, 16(3):349–364, 2008. 1, 2

[23] Jonathan Goh, Sridhar Adepu, Khurum Nazir Junejo, and Aditya Mathur. A dataset to support research in the design of secure water treatment systems. In *Critical Information Infrastructures Security*, pages 88–99. Springer International Publishing, 2017. 6

[24] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 2

[25] Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, and Nando de Freitas. Hyperbolic attention networks. In *International Conference on Learning Representations*, 2019. 3

[26] Zengyou He, Xiaofei Xu, and Shengchun Deng. Discovering cluster-based local outliers. *Pattern Recognit. Lett.*, 24:1641–1650, 2003. 2

[27] Matthias Hein, Maksym Andriushchenko, and Julian Bitterwolf. Why relu networks yield high-confidence predictions

far away from the training data and how to mitigate the problem. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 2

[28] Harold Hotelling. *The generalization of Student's ratio*. Springer, 1992. 2

[29] Anthony Hu, Zak Murez, Nikhil Mohan, Sofía Dudas, Jeffrey Hawke, Vijay Badrinarayanan, Roberto Cipolla, and Alex Kendall. FIERY: Future instance segmentation in bird's-eye view from surround monocular cameras. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2021. 2

[30] Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Soderstrom. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 387–395, 2018. 6

[31] Kyle Hundman, Valentino Constantinou, Christopher Laporte, Ian Colwell, and Tom Söderström. Detecting spacecraft anomalies using lstms and nonparametric dynamic thresholding. In *KDD*, pages 387–395. ACM, 2018. 2

[32] Borui Jiang, Ruixuan Luo, Jiayuan Mao, Tete Xiao, and Yuning Jiang. Acquisition of localization confidence for accurate object detection. In *Proceedings of the European conference on computer vision (ECCV)*, pages 784–799, 2018. 2

[33] Fabian Keller, Emmanuel Muller, and Klemens Bohm. Hics: High contrast subspaces for density-based outlier ranking. In *2012 IEEE 28th international conference on data engineering*, pages 1037–1048. IEEE, 2012. 1

[34] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? *arXiv preprint arXiv:1703.04977*, 2017. 1

[35] Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 3, 4

[36] Edwin M. Knorr and Raymond T. Ng. Algorithms for mining distance-based outliers in large datasets. In *VLDB*, pages 392–403. Morgan Kaufmann, 1998. 2

[37] Fabian Kuppers, Jan Kronenberger, Amirhossein Shantia, and Anselm Haselhoff. Multivariate confidence calibration for object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 326–327, 2020. 2

[38] Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive uncertainty estimation using deep ensembles. *arXiv preprint arXiv:1612.01474*, 2016. 1

[39] Alexander Lavin and Subutai Ahmad. Evaluating real-time anomaly detection algorithms–the numenta anomaly benchmark. In *2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, pages 38–44. IEEE, 2015. 2

[40] Aleksandar Lazarevic and Vipin Kumar. Feature bagging for outlier detection. In *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 157–166. ACM, 2005. 1

[41] John M Lee. *Riemannian Manifolds: An Introduction to Curvature*, volume 176. Springer Science & Business Media, 2006. 4

[42] John M Lee. *Smooth manifolds. In Introduction to Smooth Manifolds*. Springer, 2013. 4

[43] Dan Li, Dacheng Chen, Lei Shi, Baihong Jin, Jonathan Goh, and See-Kiong Ng. Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks. In *ICANN*, 2019. 3

[44] Longyuan Li, Junchi Yan, Haiyang Wang, and Yaohui Jin. Anomaly detection of time series with smoothness-inducing sequential variational auto-encoder. *IEEE transactions on neural networks and learning systems*, 32(3):1177–1191, 2020. 3

[45] Zhizhong Li and Derek Hoiem. Improving confidence estimates for unfamiliar examples. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2683–2692, 2020. 2

[46] Zhihan Li, Youjian Zhao, Jiaqi Han, Ya Su, Rui Jiao, Xidao Wen, and Dan Pei. Multivariate time series anomaly detection and interpretation using hierarchical inter-metric and temporal embedding. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 3220–3230, 2021. 3

[47] Qi Liu, Maximilian Nickel, and Douwe Kiela. Hyperbolic graph neural networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. 3

[48] Shaoteng Liu, Jingjing Chen, Liangming Pan, Chong-Wah Ngo, Tat-Seng Chua, and Yu-Gang Jiang. Hyperbolic visual embedding learning for zero-shot recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2020. 3

[49] Teng Long, Pascal Mettes, Heng Tao Shen, and Cees G. M. Snoek. Searching for actions on the hyperbole. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1138–1147, 2020. 3

[50] Aaron Lou, Isay Katsman, Qingxuan Jiang, Serge J. Belongie, Ser-Nam Lim, and Christopher De Sa. Differentiating through the fréchet mean. *ArXiv*, abs/2003.00335, 2020. 3

[51] Marco Maggipinto, Chiara Masiero, Alessandro Beghi, and Gian Antonio Susto. A convolutional autoencoder approach for feature extraction in virtual metrology. *Procedia Manufacturing*, 17:126–133, 2018. 7

[52] Pankaj Malhotra, Anusha Ramakrishnan, Gaurangi Anand, Lovekesh Vig, Puneet Agarwal, and Gautam M. Shroff. Lstm-based encoder-decoder for multi-sensor anomaly detection. *ArXiv*, abs/1607.00148, 2016. 3, 8

[53] Aditya P. Mathur and Nils Ole Tippenhauer. Swat: a water treatment testbed for research and training on ics security. *2016 International Workshop on Cyber-physical Systems for Smart Water Networks (CySWater)*, pages 31–36, 2016. 2, 6

[54] Lukas Neumann, Andrew Zisserman, and Andrea Vedaldi. Relaxed softmax: Efficient confidence auto-calibration for safe pedestrian detection. 2018. 2

[55] Spiros Papadimitriou, Hiroyuki Kitagawa, Phillip B Gibbons, and Christos Faloutsos. Loci: Fast outlier detection using the local correlation integral. In *Proceedings 19th International Conference on Data Engineering (Cat. No. 03CH37405)*, pages 315–326. IEEE, 2003. 1, 2

[56] Alaa Sagheer and Mostafa Kotb. Unsupervised pre-training of a deep lstm-based stacked autoencoder for multivariate time series forecasting problems. *Scientific reports*, 9(1):1–16, 2019. 7

[57] Ruslan Salakhutdinov and Geoffrey Hinton. Deep boltzmann machines. In *Artificial intelligence and statistics*, pages 448–455. PMLR, 2009. 2

[58] Enver Sangineto, Moin Nabi, Dubravko Culibrk, and Nicu Sebe. Self paced deep learning for weakly supervised object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(3):712–725, 2019. 2

[59] Hamed Sarvari, Carlotta Domeniconi, Bardh Prenkaj, and Giovanni Stilo. Unsupervised boosting-based autoencoder ensembles for outlier detection. In *PAKDD (1)*, pages 91–103. Springer, 2021. 1, 3

[60] Ryohei Shimizu, YUSUKE Mukuta, and Tatsuya Harada. Hyperbolic neural networks++. In *International Conference on Learning Representations*, 2021. 3, 5

[61] Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, page 2828–2837. ACM, 2019. 2, 8

[62] Ya Su, Youjian Zhao, Chenhao Niu, Rong Liu, Wei Sun, and Dan Pei. Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 2828–2837, 2019. 2

[63] Dídac Surís, Ruoshi Liu, and Carl Vondrick. Learning the predictability of the future. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12607–12617, 2021. 2, 3, 4, 6, 7

[64] Alexander Vakhitov, Luis Ferraz, Antonio Agudo, and Francesc Moreno-Noguer. Uncertainty-aware camera pose estimation from points and lines. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4659–4668, 2021. 2

[65] Zhiwei Wang, Zhengzhang Chen, Jingchao Ni, Hui Liu, Haifeng Chen, and Jiliang Tang. Multi-scale one-class recurrent neural networks for discrete event sequence anomaly detection. In *The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 3726–3734. ACM, 2021. 2

[66] Jianwen Xie, Yang Lu, Ruiqi Gao, Song-Chun Zhu, and Ying Nian Wu. Cooperative training of descriptor and generator networks. *IEEE transactions on pattern analysis and machine intelligence*, 42(1):27–45, 2018. 2

[67] Jianwen Xie, Yang Lu, Song-Chun Zhu, and Yingnian Wu. A theory of generative convnet. In *International Conference on Machine Learning*, pages 2635–2644. PMLR, 2016. 2

[68] Jing Zhang, Yuchao Dai, Mochu Xiang, Deng-Ping Fan, Peyman Moghadam, Mingyi He, Christian Walder, Kaihao Zhang, Mehrtash Harandi, and Nick Barnes. Dense uncertainty estimation. *arXiv preprint arXiv:2110.06427*, 2021. 2

[69] Bin Zhou, Shenghua Liu, Bryan Hooi, Xueqi Cheng, and Jing Ye. Beatgan: Anomalous rhythm detection using adversarially generated time series. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pages 4433–4439. International Joint Conferences on Artificial Intelligence Organization, 7 2019. 3

[70] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2242–2251, 2017. 3

[71] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International conference on learning representations*, 2018. 8