

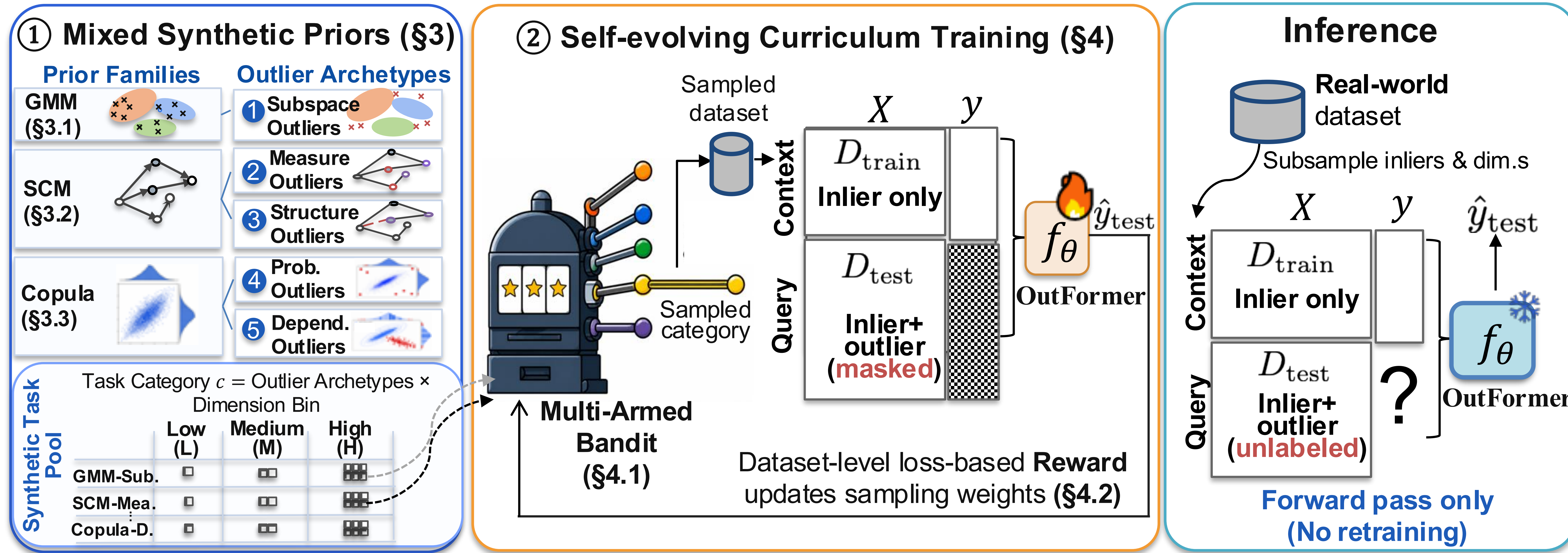


# From Zero to Hero: Advancing Zero-Shot Foundation Models for Tabular Outlier Detection

Xueying Ding<sup>1</sup>, Haomin Wen<sup>1</sup>, Simon Klüttermann<sup>2</sup>, Leman Akoglu<sup>1</sup>  
1: Carnegie Mellon University 2: Technical University Dortmund, Germany

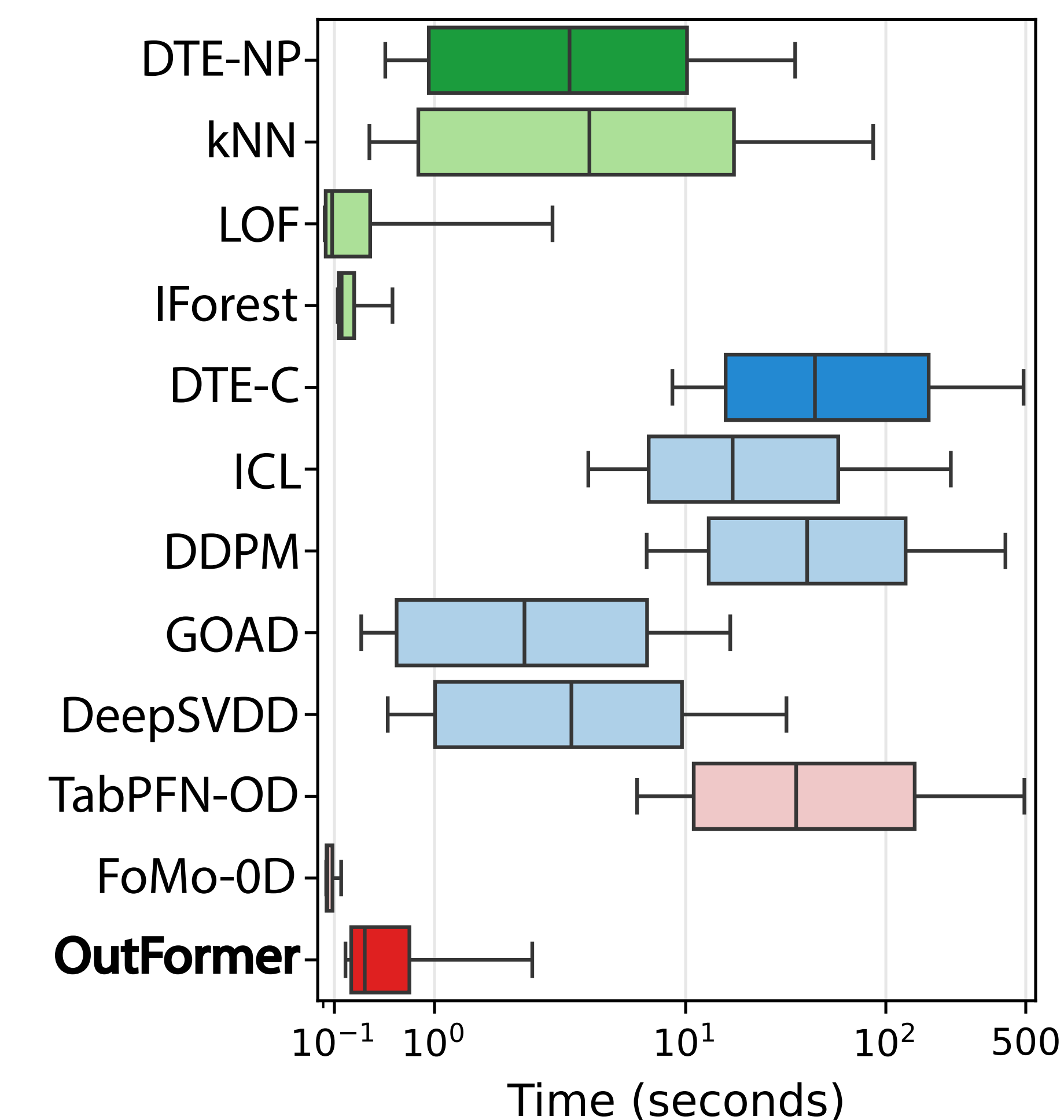
**Problem Statement:** Can we build a zero-shot foundation model for tabular outlier detection (OD) that generalizes across datasets without retraining?

## Method Overview



## Results & Time Analysis

Model	Avg. Rank (↓)	ELO (↑)	Winrate (↑)	rAUC (↑)	$C_{\Delta}$ (↓)	Win/Tie	p-val.
DTE-NP	<b>5.04</b> ±2.9	1100	<b>0.61</b>	<b>0.905</b> ±0.13	0.35	<b>0.43/0.16</b>	0.63
kNN	5.41±2.9	1129	0.57	0.899±0.13	0.35	<b>0.45/0.14</b>	0.09
LOF	6.17±3.4	864	0.51	0.869±0.15	0.42	<b>0.53/0.12</b>	<b>0.00</b>
IForest	6.57±3.6	917	0.49	0.883±0.12	0.43	<b>0.55/0.07</b>	<b>0.00</b>
DTE-C	5.94±3.2	<b>1167</b>	0.38	0.885±0.14	0.37	<b>0.51/0.14</b>	<b>0.00</b>
ICL	6.40±3.3	1000	0.49	0.871±0.14	0.39	<b>0.54/0.13</b>	<b>0.00</b>
DDPM	9.18±2.6	770	0.24	0.796±0.14	0.53	<b>0.77/0.08</b>	<b>0.00</b>
GOAD	8.72±3.6	696	0.28	0.752±0.23	0.51	<b>0.71/0.09</b>	<b>0.00</b>
DeepSVDD	6.34±3.2	994	0.49	0.869±0.15	0.40	<b>0.55/0.15</b>	<b>0.00</b>
TabPFN-OD	5.39±3.4	1141	0.58	0.901±0.13	<b>0.34</b>	<b>0.45/0.15</b>	0.21
FoMo-0D	6.45±3.9	1012	0.49	0.869±0.15	0.39	<b>0.55/0.13</b>	<b>0.00</b>
OUTFORMER	<b>5.06</b> ±3.3	<b>1209</b>	<b>0.60</b>	<b>0.903</b> ±0.12	<b>0.32</b>	-	-



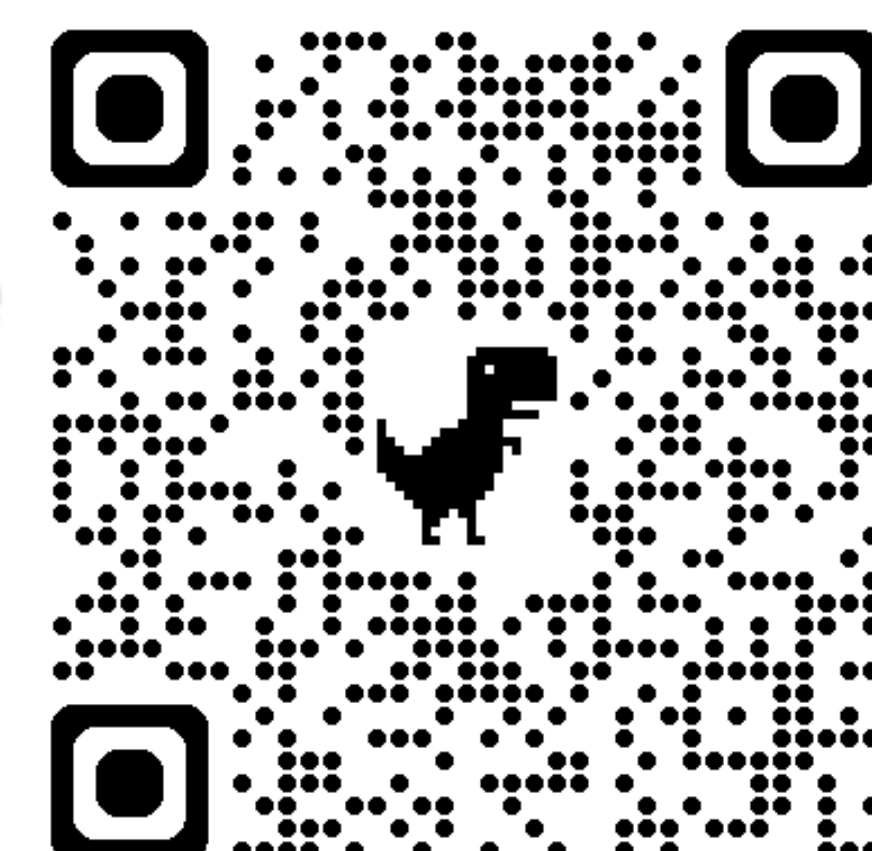
↑ Compared on ALL 1500+ datasets from all 3 OD benchmarks, OUTFORMER performs on par with the SOTA.

← OUTFORMER maintains speedy inference while achieving competitive performance.

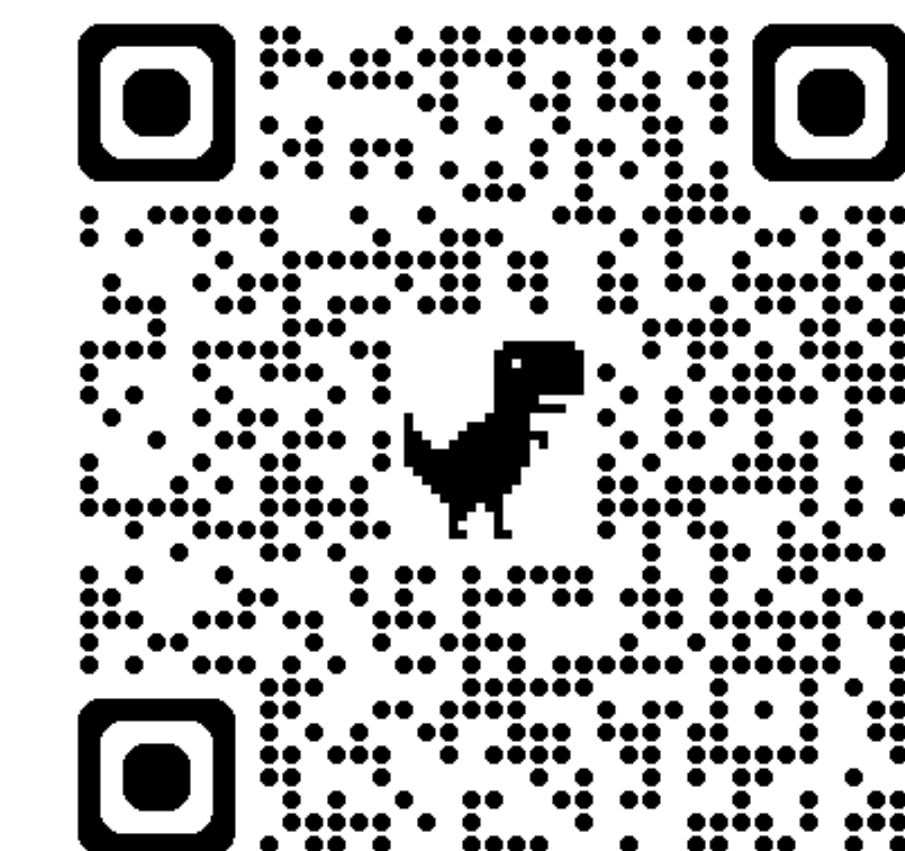
## Our Contributions

- OUTFORMER:** a fast, zero-shot foundation model for tabular outlier detection.
- Synthetic-prior pretraining:** diverse synthetic priors to simulate outliers.
- Self-evolving curriculum training (SEC):** adaptive data sampling strategy to improve foundation model pre-training.
- NEW Large-scale OD benchmarks:** over 1,500 real-world datasets and 800+ synthetic datasets for outlier detection evaluation.

Please follow us!



Paper



Dataset & code