ChatGPT

Related Work / Background for *ScalorVit-EBM*

1. Unsupervised Time-Series Anomaly Detection

1.1 Classical and Statistical Methods

- **Support Vector Method for Novelty Detection** (Schölkopf *et al.*, NIPS 1999) – Introduced the one-class SVM, a kernel-based novelty detector that finds a boundary enclosing most data [1]. It is foundational for unsupervised anomaly detection, though it can be sensitive to high-dimensional or complex data distributions.
- **Support Vector Data Description (SVDD)** (Tax & Duin, JMLR 2004) – Learns a minimum-volume hypersphere around normal data. Like one-class SVM, it provides a principled formulation for outlier scoring but also struggles as dimensionality grows.
- **Isolation Forest** (Liu *et al.*, ICDM 2008) – Proposes an ensemble method that isolates anomalies by randomly partitioning data. It operates in linear time and is reported to work well even in high-dimensional settings [2]. However, it treats each point independently and ignores any temporal or sequential structure.
- **ARIMA Models** (Moayedi & Masnadi-Shirazi, IEEE 2008) – Early work applied ARIMA forecasting to network traffic anomaly detection. ARIMA models capture linear temporal correlations but require stationarity and manual hyper-parameter tuning [3], making them brittle on complex, non-stationary series.
- **Holt-Winters (Triple Exponential Smoothing)** – A classical seasonal forecasting method; it implicitly assumes relatively stable seasonal patterns and is not typically used for high-dimensional or highly non-stationary data.
- **Summary:** Classical methods like ARIMA/Holt-Winters (linear forecasting) and shallow detectors (one-class SVM/SVDD, isolation forest) are efficient for simple or low-dimensional series, but they generally do not capture complex temporal dependencies. Recent surveys note that these approaches degrade on high-dimensional, non-stationary time series [4] [5], motivating deep learning alternatives.

1.2 Deep Learning-Based TSAD – Reconstruction Models

- **EncDec-AD (LSTM Encoder-Decoder)** (Malhotra *et al.*, ICML Workshop 2016) – Uses a stacked LSTM autoencoder to learn normal sequence reconstructions and flags anomalies by large reconstruction errors [6]. This work is often cited as a pioneer of RNN-based reconstruction AD.
- **USAD (UnSupervised Anomaly Detection)** (Audibert *et al.*, KDD 2020) – Introduces an adversarially-trained autoencoder framework (two alternating autoencoders) to improve anomaly detection stability and speed [7]. USAD reports fast convergence and strong detection on multivariate sensor data.
- **PatchTrAD (Patch-wise Transformer AD)** (Vilhes *et al.*, NeurIPS 2025) – Proposes a patch-based Transformer encoder that reconstructs segments of the time series under a reconstruction error criterion. It achieves state-of-the-art accuracy while remaining efficient at inference time [8].

- **TranAD (Transformer AD)** (Tuli *et al.*, PVLDB 2022) – Uses a self-attention (Transformer) encoder-decoder with adversarial (one-class classification) training to detect anomalies. TranAD achieves high detection performance with fast training compared to RNN baselines [9] .
- **Summary:** These methods all build deep sequence-to-sequence models (LSTM, GRU, or Transformer autoencoders) to reconstruct normal behavior. The reconstruction error serves as the anomaly score. Each advances over earlier RNNs (e.g. by using adversarial training or transformer architectures) to improve detection accuracy and training efficiency.

1.3 Deep Learning-Based TSAD – Forecasting Models

- **DeepAnT (CNN Predictor)** (Munir *et al.*, IEEE Access 2019) – Uses a 1D-CNN to predict the next time step from a sliding window; points with large forecast error are flagged as anomalies [10] . DeepAnT explicitly targets contextual and point anomalies in time series and shows strong performance on industrial benchmarks.
- **OmniAnomaly (Stochastic RNN)** (Su *et al.*, KDD 2019) – Introduces a GRU-VAE (stochastic RNN) model that learns a probabilistic representation of multivariate series. It reconstructs inputs via a variational network (with normalizing flows for flexibility) and uses low reconstruction likelihood to detect anomalies [11] . OmniAnomaly is notable for its robustness and interpretability in multivariate settings.
- **Summary:** Forecasting-based approaches train a predictive model (often RNN or CNN) to forecast future values. Anomalies are identified by abnormally large prediction errors. These methods capture sequence patterns implicitly through prediction, complementing reconstruction-based approaches.

1.4 Deep Learning-Based TSAD – Probabilistic/Distribution Models (Excluding EBMs)

- **DAGMM (Deep Autoencoding Gaussian Mixture Model)** (Zong *et al.*, ICML 2018) – Jointly trains a deep autoencoder and a Gaussian Mixture Model on its latent outputs [12] . This end-to-end model estimates the data density and uses low density regions to detect anomalies. DAGMM was one of the first to combine deep representation learning with explicit density estimation for TSAD.
- **TadGAN (Time-series Anomaly Detection with GANs)** (Geiger *et al.*, IEEE BigData 2021) – Builds on GANs by using LSTM-based generator and discriminator with cycle-consistency losses. It computes anomaly scores from a combination of reconstruction error and the critic's output [13] . TadGAN shows improved detection (highest average F1) on a range of benchmark datasets.
- *Other density models:* Methods like normalizing flows (e.g. Masked Autoregressive Flows) have been explored for novelty detection, as they provide exact likelihoods. GAN-based generative models (beyond TadGAN) have been widely studied for TSAD. These approaches contrast with EBMs by either learning an explicit likelihood (flows) or an implicit mapping (GANs).

2. Signal-to-Image Transformation for Time-Series Analysis

2.1 Time-Domain vs. Frequency-Domain

- **Trade-offs:** Time-domain analysis (e.g. raw signals or autoregressive models) directly captures temporal patterns, but may miss periodicity. Frequency-domain analysis (Fourier) reveals global spectral content but **assumes stationarity** and loses temporal localization [14] . For non-stationary signals, STFT can localize in time but suffers a fixed window resolution (time–frequency trade-off)

[14]. Wavelet-based methods (CWT/DWT) adaptively capture time-frequency features and handle non-stationarity better [15] [16].

- *(Reference)* Yao *et al.* (2025) survey notes that traditional FFT/DFT is "well-suited for stationary signals" but cannot handle non-stationary series, whereas STFT adds time localization at the cost of resolution [14]. Rhif *et al.* (2019) review highlights the wavelet transform's strength in decomposing non-stationary components [15].

2.2 Time-Frequency Representations (STFT, CWT)

- **Short-Time Fourier Transform (STFT):** Applies a sliding window FFT to obtain a spectrogram. Common in many domains, but its fixed window size imposes a time–frequency resolution limit [14]. It is simple but can miss fine temporal details if the window is not chosen carefully.
- **Continuous Wavelet Transform (CWT):** Uses scalable wavelet functions to analyze signal at multiple resolutions. CWT provides a continuous time-frequency representation that adapts to both high- and low-frequency features. As noted by Rhif *et al.* (2019), CWT "has been successfully applied…to decompose the non-stationary TS into [a] time-frequency domain" [15], making it well-suited for signals with evolving frequency content.
- **Comparison:** CWT's multi-resolution nature makes it better at capturing transient or non-stationary events than STFT [15]. However, CWT is more computationally intensive and produces redundant data (due to continuous scaling) [17]. STFT remains useful for a coarse time-frequency analysis, but modern TSAD often prefers wavelets or learned transforms for non-stationary data.

2.3 Other "Time-Series as Images" Techniques

- **Gramian Angular Field (GAF) & Markov Transition Field (MTF)** (Wang & Oates, KDD 2015) – Transform a 1D series into 2D images. GAF represents temporal correlations via trigonometric sums in a Gram matrix, preserving time order along the diagonal [18]. MTF encodes transition probabilities of quantized values into a matrix. These encodings allow CNNs to be applied to time series by leveraging spatial patterns. Wang & Oates show that features learned from GAF/MTF images are informative for classification and imputation tasks [18].
- **Recurrence Plots:** (Marwan *et al*., Phys. Rep. 2007) While not a deep learning paper, recurrence plots visualize times at which a trajectory in phase-space returns to previous states. They have been used to reveal hidden patterns and, more recently, fed into CNNs for anomaly and pattern detection in time series.
- *Summary:* These techniques create image-like representations of time-series data (Gramian fields, Markov fields, recurrence matrices) that capture temporal dynamics as spatial patterns. Such images enable the use of convolutional or vision-based models for time-series tasks.

3. Energy-Based Models (EBMs) for Anomaly Detection

3.1 Fundamentals of EBMs

- **Deep Structured Energy-Based Models (DSEBM)** (Zhai *et al*., ICML 2016) – Proposes modeling the negative log-density (energy) of data directly with a deep network output. The model is trained with score matching (a variant of contrastive divergence) to avoid expensive MCMC. Zhai *et al.* connect EBMs to autoencoders, using energy and reconstruction error jointly for anomaly scoring [19]. This work illustrates the core EBM idea: learn an energy function $E(x)$ such that $p(x)\propto e^{-E(x)}$.

- **Other Theory:** Traditional EBMs (e.g., Hopfield nets, RBMs, and latent variable models) also operate by assigning low energy to observed data and high energy elsewhere. Early tutorials (LeCun *et al*., 2006; Ngiam *et al*., 2011) discuss EBMs as an expressive but often hard-to-train framework. Recent work emphasizes EBMs' principled basis (no need for latent variables) and flexibility in incorporating constraints or priors.

3.2 Advantages Over Other Generative Models

- **EBM vs. VAE/Flow:** Yoon *et al.* (NeurIPS 2023) observe that likelihood-based models (VAEs, normalizing flows) often assign high likelihood to outliers [20], undermining anomaly detection. In contrast, EBMs trained by maximum likelihood explicitly sample "negative" points via MCMC and drive their likelihood down during training. This leads to sharper boundaries around the data manifold. The authors report that deep EBMs significantly outperform VAEs and flows on standard AD benchmarks [20].
- **EBM vs. GAN:** GANs can generate realistic samples but lack an explicit likelihood. EBMs explicitly model (unnormalized) densities, making them natural for detecting low-density anomalies. Some surveys note that unlike GANs or VAEs, EBMs do not suffer from easy modes to fool (mode collapse) because the energy function directly evaluates any input.
- *Key Point:* The iterative negative-sampling training of EBMs gives them robustness to out-of-distribution data [21], providing an advantage in anomaly/outlier detection tasks.

3.3 Training EBMs (Contrastive Divergence, Flow/Energy Matching)

- **Contrastive Divergence (CD)** (Hinton, 2010) – A classical EBM training algorithm for RBMs and similar models. CD approximates the likelihood gradient by a short MCMC chain. Hinton's practical guide notes that "RBMs are usually trained using the contrastive divergence learning procedure" [22]. CD makes EBMs tractable but can suffer from slow mixing and biased updates.
- **Score Matching & Pseudo-Likelihood:** Alternatives to CD include score matching and noise-contrastive estimation, which avoid full MCMC. For example, Zhai *et al.* use a form of score matching to link an EBM with a denoising autoencoder [19].
- **Flow Matching / Energy Matching:** Recent works aim to unify EBMs with flow-based sampling. In "Energy Matching" (Balcerak *et al*., NeurIPS 2025), a single scalar energy field is learned such that samples flow (via optimal transport) towards the data manifold, where an energy term then concentrates them into a Boltzmann distribution [23]. This approach requires no auxiliary diffusion timesteps and achieves high-quality generation and density modeling. It demonstrates a new way to train EBMs by combining ideas from flows and diffusion, greatly simplifying training while improving sample fidelity [23].
- **Summary:** Early EBM training (CD, MCMC) is challenging for large models, motivating innovations like score matching and flow-based methods. Recent frameworks (e.g. "Energy Matching") remove the need for diffusion and offer a unifying theory for generative modeling with EBMs [23].

```
@inproceedings{scholkopf1999novelty,
  title={Support Vector Method for Novelty Detection},
  author={Sch{\"o}lkopf, Bernhard and Williamson, Robert C. and Smola, Alex J.
and Shawe-Taylor, John and Platt, John C.},
  booktitle={Advances in Neural Information Processing Systems},
```

```
  year={1999}
}
```

Summary: Kernel one-class SVM for novelty detection [1].

```
@article{tax2004support,
  title={Support Vector Data Description},
  author={Tax, David MJ and Duin, Robert PW},
  journal={Machine Learning Journal},
  year={2004}
}
```

Summary: SVDD learns a minimum-radius hypersphere around data.

```
@inproceedings{liu2008isolation,
  title={Isolation Forest},
  author={Liu, Fei Tony and Ting, Kai Ming and Zhou, Zhi-Hua},
  booktitle={IEEE International Conference on Data Mining (ICDM)},
  year={2008}
}
```

Summary: Random partitioning ensemble that isolates anomalies efficiently [2].

```
@inproceedings{moayedi2008arima,
  title={ARIMA Model for Network Traffic Prediction and Anomaly Detection},
  author={Moayedi, Shahram H and Masnadi-Shirazi, Mohammad},
  booktitle={International Symposium on Information Technology (IT)},
  year={2008}
}
```

Summary: Applies ARIMA forecasting to detect network anomalies; highlights ARIMA's stationarity and tuning limitations [3].

```
@inproceedings{malhotra2016lstmencdec,
  title={LSTM Encoder-Decoder for Multi-sensor Anomaly Detection},
  author={Malhotra, Pankaj and Vig, Lovekesh and Shroff, Gautam and Agarwal,
Puneet},
  booktitle={ICML Anomaly Detection Workshop},
  year={2016}
}
```

Summary: An LSTM autoencoder that reconstructs normal sequences; uses reconstruction error for unsupervised TSAD [6].

```
@inproceedings{audibert2020usad,
  title={USAD: UnSupervised Anomaly Detection on Multivariate Time Series},
  author={Audibert, Adrien and Brecheisen, Thomas and Mouaddib, Anis and
Chambres, Anne and Le Gland, Fabien and Kominiarczuk, Jarosław and Petit,
Nicolas},
  booktitle={ACM SIGKDD Conference on Knowledge Discovery and Data Mining
(KDD)},
  year={2020}
}
```

Summary: An adversarial autoencoder approach with two alternating modules, achieving fast training and strong anomaly separation [7].

```
@article{munir2018deepant,
  title={DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in
Time Series},
  author={Munir, Mohsin and Siddiqui, Shoaib Ahmed and Dengel, Andreas and
Ahmed, Sheraz},
  journal={IEEE Access},
  year={2019}
}
```

Summary: A CNN forecaster that predicts the next time step; large prediction errors indicate anomalies [10].

```
@inproceedings{su2019omnianomaly,
  title={Robust Anomaly Detection for Multivariate Time Series through
Stochastic Recurrent Neural Network},
  author={Su, Ya and Zhao, Youjian and Niu, Chenhao and Liu, Rong and Sun, Wei
and Pei, Dan},
  booktitle={ACM SIGKDD Conference on Knowledge Discovery and Data Mining
(KDD)},
  year={2019}
}
```

Summary: Introduces a GRU-VAE model (OmniAnomaly) with normalizing flows; uses reconstruction probability to detect anomalies [11].

```
@inproceedings{zong2018dagmm,
  title={Deep Autoencoding Gaussian Mixture Model for Unsupervised Anomaly
Detection},
```

```
  author={Zong, Bo and Song, Quanming and Ruan, Miao and ...

, author continuation...
  booktitle={Proceedings of the 23rd ACM SIGKDD International Conference on
Knowledge Discovery and Data Mining},
  year={2018}
}
```

Summary: Jointly trains a deep autoencoder and a Gaussian mixture on its latent outputs to estimate density; anomalies are low-probability points [12].

```
@inproceedings{geiger2021tadgan,
  title={TadGAN: Time Series Anomaly Detection Using Generative Adversarial
Networks},
  author={Geiger, Alexander and Liu, Dongyu and Alnegheimish, Sarah and Cuesta-
Infante, Alfredo and Veeramachaneni, Kalyan},
  booktitle={IEEE International Conference on Big Data},
  year={2021}
}
```

Summary: GAN-based detector using LSTM generator/critic with cycle-consistency; combines reconstruction and critic scores for anomaly detection [13].

```
@article{rhif2019wavelet,
  title={Wavelet Transform Application for/in Non-Stationary Time-Series
Analysis: A Review},
  author={Rhif, Manel and Ben Abbes, Ali and Farah, Imed Riadh and Sang,
Yanfang},
  journal={Applied Sciences},
  year={2019}
}
```

Summary: Surveys use of wavelet transforms (CWT/DWT) for non-stationary signals, noting wavelets' strength in time-frequency decomposition of complex time series [15].

```
@article{wang2015imaging,
  title={Imaging Time-Series to Improve Classification and Imputation},
  author={Wang, Zhiguang and Oates, Tim},
  journal={Neurocomputing},
  year={2015}
}
```

Summary: Introduces Gramian Angular Fields and Markov Transition Fields to encode time series as images. GAF preserves temporal dependencies (e.g. main diagonal retains original signal) [18], enabling CNN-based classification of series.

```
@inproceedings{zhai2016deep,
  title={Deep Structured Energy Based Models for Anomaly Detection},
  author={Zhai, Shuangfei and Cheng, Yu and Lu, Weining and Zhang, Zhongfei},
  booktitle={International Conference on Machine Learning (ICML)},
  year={2016}
}
```

Summary: Proposes using deep neural networks to define energy (negative log-density) for data. Trains via score matching (linking EBMs with autoencoders) to model complex distributions without expensive sampling [19].

```
@article{yoon2023manifold,
  title={Energy-Based Models for Anomaly Detection: A Manifold Diffusion
Recovery Approach},
  author={Yoon, Sangwoong and Jin, Young-Uk and Noh, Yung-Kyun and Park, Frank
C.},
  journal={arXiv preprint arXiv:2310.18677},
  year={2023}
}
```

Summary: Shows that likelihood models like VAEs/flows often fail on AD, whereas EBMs (via MCMC negatives) can more sharply delineate normal data [20]. EBMs achieve notably better OOD detection by explicitly lowering the probability of generated negatives.

```
@techreport{hinton2010rbm,
  title={A Practical Guide to Training Restricted Boltzmann Machines},
  author={Hinton, Geoffrey},
  institution={University of Toronto},
  year={2010}
}
```

Summary: Practical tips for RBMs; notes that RBMs are “usually trained using the contrastive divergence” method [22], highlighting CD as the standard EBM training procedure.

```
@inproceedings{balcerak2025energy,
  title={Energy Matching: Unifying Flow Matching and Energy-Based Models for
Generative Modeling},
  author={Balcerak, Michal and Amiranashvili, Tamaz and Terpin, Antonio and
Shit, Suprosanna and Bogensperger, Lea and Kaltenbach, Sebastian and
```

```
Koumoutsakos, Petros and Menze, Bjoern},
  booktitle={NeurIPS},
  year={2025}
}
```

Summary: Proposes a flow+energy framework where samples move from noise to data manifold via optimal transport, then an “energy” term fine-tunes them into a Boltzmann distribution [23]. This single-field model trains without diffusion steps and outperforms prior EBMs on CIFAR-10/ImageNet, advancing EBM training methodology.

---

[1] Support Vector Method for Novelty Detection
http://papers.neurips.cc/paper/1723-support-vector-method-for-novelty-detection.pdf

[2] (PDF) Isolation Forest
https://www.researchgate.net/publication/224384174_Isolation_Forest

[3] Online Forecasting and Anomaly Detection Based on the ARIMA Model
https://www.mdpi.com/2076-3417/11/7/3194

[4] Unsupervised anomaly detection in time-series: An extensive evaluation and analysis of state-of-the-art methods
https://arxiv.org/pdf/2212.03637

[5] [12] A Survey of Deep Anomaly Detection in Multivariate Time Series: Taxonomy, Applications, and Directions
https://www.mdpi.com/1424-8220/25/1/190

[6] [1607.00148] LSTM-based Encoder-Decoder for Multi-sensor Anomaly Detection
https://arxiv.org/abs/1607.00148

[7] (PDF) USAD: UnSupervised Anomaly Detection on Multivariate Time Series
https://www.researchgate.net/publication/343779877_USAD_UnSupervised_Anomaly_Detection_on_Multivariate_Time_Series

[8] [2504.08827] PatchTrAD: A Patch-Based Transformer focusing on Patch-Wise Reconstruction Error for Time Series Anomaly Detection
https://arxiv.org/abs/2504.08827

[9] vldb.org
http://vldb.org/pvldb/vol15/p1201-tuli.pdf

[10] DeepAnT: A Deep Learning Approach for Unsupervised Anomaly Detection in Time Series
https://www.dfki.de/fileadmin/user_upload/import/10175_DeepAnt.pdf

[11] Robust Anomaly Detection for Multivariate Time Series through Stochastic Recurrent Neural Network
https://netman.aiops.org/wp-content/uploads/2019/08/OmniAnomaly_camera-ready.pdf

[13] [2009.07769] TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks
https://arxiv.org/abs/2009.07769

[14] [16] [17] Beyond the Time Domain: Recent Advances on Frequency Transforms in Time Series Analysis
https://arxiv.org/html/2504.07099v1

[15] Wavelet Transform Application for/in Non-Stationary Time-Series Analysis: A Review
https://www.mdpi.com/2076-3417/9/7/1345

[18] Imaging Time-Series to Improve Classification and Imputation
https://arxiv.org/pdf/1506.00327

[19] Deep Structured Energy Based Models for Anomaly Detection
http://proceedings.mlr.press/v48/zhai16.pdf

[20] [21] arxiv.org
https://arxiv.org/pdf/2310.18677

[22] cs.toronto.edu
https://www.cs.toronto.edu/~hinton/absps/guideTR.pdf

[23] [2504.10612] Energy Matching: Unifying Flow Matching and Energy-Based Models for Generative Modeling
https://arxiv.org/abs/2504.10612