

Representation Learning for Zero-Shot Anomaly Detection

Johannes Hölker

¹ University of Kassel

² Germany

johannes.hoelker@uni-kassel.de

Abstract. The abstract should briefly summarize the contents of the paper in 150–250 words.

Keywords: Representation Learning · Zero Shot Learning · Anomaly Detection.

1 Introduction

At the time this paper is written sensors can be found everywhere and their usage is increasing in many fields. Gyroscopes, cameras, compasses and accelerometers are integrated in smartphones. Physical machines are tracking their movement through vibration sensors, health care systems in hospitals visualize the heart beat of a patient and voltmeters measure the generated power in a solar plant. Everytime sensor values are collected over time and time series data (TSD) is produced.

In some scenarios the measurements of different sensors are combined. Physical machines sometimes track vibration and motor rotations, health care systems visualize heart beat and blood sugar level and solar plants measure voltage and current. Collections of different sensors measuring at a common time window produce Multi Variate Time Series Data (MVTSD).

Applications that produce MV-TSD evaluate the data and further decisions and actions depend on a correct analysis. Normally the data is consistent and values change constantly in repetitive patterns. This is when the machine, the patient health or the solar plant is functioning like it is supposed to. But sometimes the values change unpredicted because of differing surroundings or other influences. This can lead to serious situation, especially in the health scenario. When the patient's heart beat changes its pattern the health of the patient is endangered. Machine measurements can detect a fault and a solar plant may detect a decline in the generated power.

Recognising and reacting to these changes in MVTSD can therefore be very important. But these interruptions occur in different forms. They can be recognised as outliers or they are hidden and not obviously seen as anomalies. In some cases they form shapes which never occurred before. This raises the demand for a tool to detect anomalies in time series data without any further knowledge of the anomaly.

A systematic literature research concerning the topic is conducted and the best choices are implemented on a test data set. In 2 the basic methodology used in this paper is described and the main research questions are formulated. First basic necessary terms are explained in 3. In 4 the found papers and their methodologies are presented and explained. The methodologies are compared and evaluated for usability in the context of Zero-Shot Anomaly Detection in 5. The implementation of suiting techniques is provided in 6. Finally the results are discussed and concluded in 7

2 Systematic Literature Review

A literature review to contribute in the development of an anomaly detection tool is presented in this paper. It provides an overview of the latest trends in representation learning and extracts the possible solutions addressing the problem described in 1. The review conforms to the methodology presented by Kitchenham et al. (2009). First the research questions are formulated. Finally Inclusion and Exclusion Criteria are formulated in order to filter the found literature for the application. The search process and the websites used are listed.

Further analysis with a systematic quality assessment and data collection like in Kitchenham et al. (2009) are excluded.

2.1 Research question

The topic covered includes different areas of machine learning, all being further developed in recent years. In order to break down the topic the following research questions are formulated:

- RQ1: What types of representation learning exist?
- RQ2: Which RL types can be used for multi variate time series?
- RQ3: Which of the recent methodologies can be used for anomaly detection?
- RQ4: Are the methods useful for Zero Shot Learning Scenarios?

These question form a path for further chapters. They build the basis for the formulation of the following Criteria.

2.2 Inclusion and Exclusion Criteria

This paper focuses on published methods for anomaly detection in Zero Shot Scenarios on multi variate time series data. In order to structure the search for and selection of relevant articles, the necessary guidelines are formulated below. Articles that are to be considered in more detail must meet the following inclusion criteria:

- IC1: Is the method using a representation learning concept. The focus is on methods learning representations in data using machine learning concepts.
- IC2: Does the method handle time series data?

- IC3: Is the method used for Anomaly Detection. Sometimes the proposed methods perform well on different use cases. If one of them is Anomaly Detection, the paper is included in the review.
- IC4: Published in recent years (< 6 years)

The chosen articles are examined in more detail. They are described and explained in 4. Using the gained knowledge all described articles are filtered by the following exclusion criteria in 5.

- EC1: Is the method tested on Zero-Shot Learning?
- EC2: Can the method handle multiple input variables?
- EC3: Multiple publications reporting the same methodologies. Only the latest and most promising methods are used.
- EC4: Open Source Availability. Is the method publicly available and does it contain a description on how to implement and reproduce the outcomes.

2.3 Search process

A manual search of specific conference proceedings and journal papers was made. Considering the pace on which new developments emerge in the area of machine learning the help of research tools was needed. Specifically in the field of anomaly detection the publications are made in recent years. This makes it difficult to assure finding every relevant paper.

The main tool used to find papers was Consensus, which is an academic search engine. They use large language models (LLMs) and purpose-built search technology. The chatbot is based on ChatGPT 4.0 and it should answer questions based on papers including their reference. For reassuring the existence of the paper the university bibliography KARLA is used to download and read the papers manually.

3 Definitions and Conventions

The basic expressions used in this paper are explained in the following chapter. First Representation Learning is defined and the different approaches to find representations in data are explained. A description on how to evaluate the found representations is given. Afterwards Zero Shot Learning as well as Anomaly Detection is described and explained.

3.1 Representation Learning

Representation Learning (RL) tries to detect meaningful interconnections in data relevant for further data analysis. These interconnections represent abstract information, so called background knowledge ($?, ?$).

In neural networks representations are learned in every layer. Considering hidden layers representations are hardly comprehensible for humans. They are stored in weights and biases and build so called neural representations. The

next abstraction level are spatial representations which are represented as points or regions in conceptual spaces. Transforming the information from conceptual spaces into words forms symbolic representations. The three types build cognitive representations (Gärdenfors, 2000).

To extract symbolic or spatial features which are more comprehensible for humans a knowledge discovery process with different methods of machine learning and data mining methods are used. RL techniques are divided into Propositionalization as symbolic representations and Embeddings as numeric representations (Gärdenfors, 2000, p. 4).

RL occurs in several machine learning areas. Depending on the underlying concept different strategies to extract representations can be found. They work different in detecting patterns and store them in different ways (Bishop, 2006).

In the book of Goodfellow, Bengio, and Courville (2016, p. 525) a general detailed description of representation learning is given. They summarize that representations should make the subsequent learning tasks easier. This implies that to find the best fitting representation and the underlying representation learning technique, we need to know the task it should perform afterwards.

Concepts The most straight-forward approach to detect representations are Multi Layer Perceptrons (MLP). A input vector is processed by artificial neurons passed through several layers of neurons, interconnected between each layer. The adjustment of the interconnections using weights and biases enables a learning process where labeled data is learned using backpropagation. The learns a training data set and is used to label unknown data afterwards (Nielsen, 2015)

Convolutional Neural Networks (CNN) are a variation of MLP building subsets of the input vector. This is mainly used in image processing.

Recurrent neural networks (RNN) are another variation of the tradition feed-forward MLP. Every neuron has a additional input containing the previous state. This is useful for time series data.

One solution to learn representations is Contrastive Learning (CL). Pairs of data points are labeled as similar and dissimilar. These data points are put into a feature space where the distance between the two represents their similarity. With a contrastive loss function and a label of similarity between two points, the model is trained by putting the similar data points together and separating dissimilar points. Using this method groups of similar data points are formed.

Autoencoders are another important method in representation learning. An autoencoder is a type of neural network used to learn efficient codings of input data in an unsupervised manner. It consists of an encoder that compresses the input into a latent-space representation and a decoder that reconstructs the input from this representation. The goal is to minimize the difference between the input and the reconstructed output.

Transformers, initially developed for natural language processing tasks, have become a powerful tool in representation learning. They use self-attention mechanisms to weigh the significance of each part of the input data differently, enabling the model to capture long-range dependencies (Vaswani et al., 2017).

In summary, representation learning can be achieved using different techniques, each suitable for different types of data and tasks. From neural networks and autoencoders to transformers, these methods provide the tools necessary to transform raw data into meaningful representations that facilitate further analysis and learning.

Evaluation This chapter describes how to evaluate the performance of a RL approach. ? (?) describe what makes a representation "good". They list the following factors:

- Smoothness
- Multiple Explanatory Factors
- A hierarchical organization of explanatory factors
- Semi-supervised learning
- Shared factors across tasks
- Manifolds
- Natural clustering
- Temporal and spatial coherence
- Sparsity
- Simplicity of factor dependencies

"We want to find properties of the data but at the same time we don't want to loose information about the input" (Goodfellow et al., 2016, S. 525)

3.2 Zero Shot Learning

Zero Shot Learning is an extreme form of transfer learning (Goodfellow et al., 2016, S. 536). While transfer learning is the concept of transferring the knowledge and weights gained at one task using them at solving another task, Zero-Shot Learning means there are no samples for the other task. The transformation of knowledge can help solving tasks where there are few or no samples available. The gained knowledge is normally stored as representations in data. Representations which are abstract enough to not see a specific item but information about items which can be applicated to groups of items. This also means that Zero-Shot Learning is only possible because additional information has been discovered during training (Goodfellow et al., 2016, S. 536).

Palatucci, Pomerleau, Hinton, and Mitchell (2009) were the first to implement a successful Zero-Shot Detection followed by Socher et al. (2013) who used semantic word vector representations to classify words in groups and to sort new words with an accuracy of 90% with a fully unsupervised model.

3.3 Anomaly Detection

Several definitions of anomalies in data can be found in literature. In this paper the definition of Gruhl (2022, S. 54) is used. It seperates anomaly and novelty detection as different tasks. Anomalies can be understood as outliers from the

regular class. But these anomalies can vary in their cause. If there is a specific cause and the anomalies occur in its own cluster, they form a novelty. If instead the outliers randomly occur with no specific root cause, they are called noise. The cause for noise then is of a different kind and cannot be classified.

Instead of dividing anomalies by their cause the shape of anomalies can vary in several ways. In real measurement data any shape is possible and it is totally unpredictable (?, ?). For training purposes anomaly injection is crucial. Then the anomalies are simulated as point anomalies or subsequence anomalies. Point anomalies occur once and can be global or contextual. Subsequence anomalies on the other hand change the values in a given time window or on long term. They can be divided in seasonal, shapelet and trend anomalies. Seasonal and shapelet change the values in a limited time window, trend anomalies are changing all following values (Darban, Webb, Pan, Aggarwal, & Salehi, 2024, p. 9).

In this paper we want to focus on single time events, which are in any case anomalies. Potentially being caused by an unknown process, they cannot be classified (Gruhl, 2022). This defines our goal as an anomaly detection task.

4 Representation Learning Methods

In this chapter the found literature is put into context. Starting with classical literature about the fundamental findings followed by actual trends in the Area of Representation Learning with different RL Strategies listed and compared. Any paper proposing a representation learning strategy used for time series data with adaptability on anomaly detection tasks is presented. Here the inclusion criteria as described in 2.2 are used to filter the found literature.

Variations in data are not always visible for a human and even less possible to label them accordingly. Like (?, ?) mentioned it is important for artificial intelligence to detect these representations in data by machines. A machine should be able to extract information hidden in the low-level sensor measurings and continue working with the representations instead of the raw data. This is according to the paper the main requirement for a good representation, to be able using it as an input to a supervised predictor.

Since the paper came out in 2013, several representation learning techniques were developed and some of them are directly applicable for time series data. In (Q. Sun & Ge, 2021) the importance of machine learning in sensor data is emphasized. They summarize deep learning techniques on data-driven soft-sensors. Soft-sensors represent hard to measure variables by adapting available sensor data.

The different RL strategies are listed, explained and compared. The strategies are organized by their underlying concept. We begin with straight-forward methods which are based on one concept and increase the complexity throughout the chapter. In the end methods which use combinations of different concepts are presented.

MLP Using a simple Multi Layer Perceptron (MLP) is a straight-forward way to learn representations and to detect anomalies in time series data (Nielsen, 2015). The input variable for the MLP are time points and the output variable represents the value at these time points. The model is trained to learn this mapping. With the trained model, the values in a live scenario are predicted and the difference to the actual values is measured. If this representation error exceeds a certain threshold, an anomaly is found (Jeong & Shin, 2022). The method is called INRAD, Implicit Neural Representation of Time-Series Data. The model is trained with data including anomalies so it is not suitable for Zero-Shot Learning.

RNN Y. Su et al. (2019) propose a method called OmniAnomaly for anomaly detection in multivariate time series data using a Stochastic Recurrent Neural Network (SRNN). This approach addresses the challenge of detecting anomalies in complex, high-dimensional time series data. Their method utilizes a SRNN to model the temporal dependencies in multivariate time series data. By incorporating stochastic units into the recurrent neural network, the model can capture the underlying uncertainty and variability in the data. This allows for more accurate detection of anomalies, as the model can differentiate between normal fluctuations and genuine anomalies. The key advantage of this method is its robustness to noisy and high-dimensional data. The SRNN learns to represent the normal patterns in the time series and identifies deviations from these patterns as anomalies (Y. Su et al., 2019). It relies on training with data that contains normal patterns, which the model uses to detect anomalies based on deviations from these patterns. Zero-shot learning requires the ability to detect anomalies without having seen any examples during training. Since OmniAnomaly depends on having access to representative normal data to learn patterns, it is not suitable for zero-shot scenarios where the model needs to generalize to completely unseen anomalies.

CNN Methods based on Convolutional Neural Networks (CNN) are normally used to classify images. In recent papers they are used to detect anomalies in images. Aota, Tong, and Okatani (2023) develop a Texture Anomaly Detection and achieve a high performance in Zero Shot Learning. They compare Zero-Shot against Many-Shot Learning in their work. Several image anomaly detection tools can be found ((Sabokrou, 2018), (Aota et al., 2023)). But CNNs perform on time series data as well.

The main idea of using CNNs is to predict a value based on the input frame. If the distance between the predicted and the actual value exceeds a predefined threshold, the anomaly can be detected. This idea is used to detect cyberattacks in industrial control systems. The study by Kravchik and Shabtai (2018) uses a dataset from a Secure Water Treatment testbed to identify cyber anomalies by measuring the statistical deviation between predicted and observed values. They explore different deep learning architectures, including CNNs and recurrent networks, and find that 1D CNNs perform particularly well for time series

prediction tasks. Unlike traditional recurrent networks, CNNs are shown to be faster, smaller, and more efficient while maintaining high detection accuracy. Their approach successfully detects the majority of cyber attacks with minimal false positives, highlighting the effectiveness of CNNs in real-time anomaly detection (Kravchik & Shabtai, 2018). However, the paper does not discuss the usability on zero-shot learning. In the same area a method detecting unknown cyber-attacks is presented by Zhang, Liu, Qiu, Zhou, and Zhang (2020) who use an Autoencoder which is discussed later on.

He and Zhao (2019) use Temporal Convolutional Networks. The core idea of the paper is to utilize TCNs, which are well-suited for sequential data due to their ability to capture temporal dependencies, to detect anomalies in time series. By training on normal patterns, the network learns to predict future values. Significant deviations between these predictions and actual observations indicate potential anomalies. The inclusion of a multivariate Gaussian model for error handling and the multi-scale feature mixture method enhances the robustness and accuracy of the anomaly detection process.

CL Learning representations in time series data is done in several different ways. One solution according to ? (?) is debiased contrastive learning. By comparing pairs of data points and rating the similarities as distances between the two, contrastive learning gets less dependent on labeled data. The data can be more general and the extracted representations are more robust. The pairs of data points are labeled as positive and negative pairs with a distance according to their similarities. With this distance they are put into a feature space where they form groups of data points. To minimize the bias between representations multigranularity augmented view generation and expert knowledge are used during training. The proposed framework is applied on industrial fault detection. The two data sets consist of various vibration signals of industrial machines and stiction sensors. The effectiveness of the proposed framework is demonstrated through its application to these datasets, where it shows improved performance in fault detection compared to traditional methods (?, ?).

Contrastive Representation Learning is also used for anomaly detection in time series data by Darban et al. (2024). They use CL combined with synthetic anomaly injection. CL enables them to capture patterns in time series data and the framework shows good results on common real world datasets. Similar to the previous paper, dissimilar pairs, the anomalies, build distant data points and similar data points are close to each other. In order to train the model artificial anomalies are injected which build distant pairs. In the next stage the classification is done by the proximity of the neighbours in the representation space. Additionally anchor points representing the nearest and furthest neighbour are given from each representation. Their methodology is called CARLA and is not tested for Zero-Shot Learning.

The article by Ngu and Lee (2023) introduces CL-TAD, a novel method for time series anomaly detection that leverages contrastive learning and reconstruction-based techniques to address the challenges of temporal dynam-

ics, label scarcity, and data diversity in real-world applications. The method comprises two main components: positive sample generation and contrastive-learning-based representation learning. Positive samples are generated by reconstructing masked parts of the time series data, helping the model learn the underlying normal patterns. These samples, along with the original data, are then fed into a contrastive learning framework, which contrasts pairs of similar (positive) and dissimilar (negative) samples to learn representations. This process helps the model map similar data points closer together in the feature space while distancing dissimilar points, making it easier to detect deviations indicative of anomalies. Extensive experiments on nine benchmark datasets show that CL-TAD outperforms ten other recent methods in detecting anomalies, highlighting its effectiveness in handling diverse and complex time series data (Ngu & Lee, 2023). While CL-TAD is not explicitly designed as a zero-shot learning method, its use of contrastive learning and reconstruction-based techniques suggests that it could have potential in zero-shot anomaly detection scenarios. However, this would depend on the model’s ability to generalize from the learned normal patterns to detect unseen anomalies. Further empirical studies would be needed to validate its performance in zero-shot learning scenarios.

To succeed on Zero-Shot Anomaly Detection, the method of One-Class Classification (OCC) can solve the problem. By gathering all "normal" values into a single class the outliers are directly detected if they are outside of it. The COCA (Contrastive One-Class Anomaly Detection) method combines contrastive learning with OCC to improve anomaly detection in time-series data. By treating original and reconstructed representations as positive pairs, it optimizes a contrastive one-class loss function that enhances the detection of anomalies while preventing common issues like hypersphere collapse. Although COCA is designed for self-supervised anomaly detection, its ability to learn from unlabeled data suggests potential applicability in zero-shot learning scenarios, though this has not been explicitly tested (Wang et al., 2023).

The paper by Lee, Byun, and Baek (2023) presents an approach for detecting anomalies also using OCC in industrial time series data, which typically lacks labels for supervised learning. They combine OCC with contrastive learning to define a new objective function that can simultaneously learn from both models. This method enhances feature extraction while preserving temporal characteristics. The paper demonstrates the method’s effectiveness through high anomaly detection performance on datasets with similar normal and anomalous data forms, highlighting its potential in industrial applications (Lee et al., 2023).

Unlike traditional OCC methods that map all normal instances into a single hypersphere, the method presented by Chen, Feng, and Wirjanto (2023) focuses on local contextual information. By pulling each normal instance towards its recent context window, it aims to better detect context-based anomalies. To prevent representation collapse, the model incorporates a deterministic contrastive loss, which improves the network’s ability to distinguish between normal and abnormal data.

Yue et al. (2022) introduce TS2Vec, a framework for learning robust and universal time series representations at multiple semantic levels through hierarchical contrastive learning. This approach utilizes timestamp masking and random cropping to create augmented context views, enhancing position-agnostic and comprehensive representations. By combining instance-wise and temporal contrastive losses, TS2Vec captures unique characteristics of different time series instances and dynamic temporal patterns within each series. Extensive experiments show that TS2Vec outperforms state-of-the-art methods in classification, forecasting, and anomaly detection tasks across 125 UCR and 29 UEA datasets, achieving average improvements of 2.4% and 3.0% in classification accuracy, respectively. The framework’s efficiency in training time further underscores its practical utility. TS2Vec demonstrates its versatility by excelling in multiple time series analysis tasks, making it a significant contribution to the field. The framework’s hierarchical contrastive learning at various scales encapsulates rich and meaningful patterns in time series data (Yue et al., 2022).

Another paper by Jiao, Yang, Song, and Tao (2022) introduces an autonomous system for anomaly detection in multivariate time series data using Contrastive Learning. The proposed TimeAutoAD automates model configuration and hyperparameter optimization, addressing challenges such as limited labeled anomaly data. It uses self-supervised contrastive learning to enhance the model’s ability to differentiate normal and anomalous time series by generating pseudo-negative samples. The method is tested on real-world datasets, demonstrating improved performance over existing anomaly detection techniques, especially in scenarios where training data may be contaminated (Jiao et al., 2022).

The paper does not explicitly address zero-shot anomaly detection, which refers to detecting anomalies without having seen any labeled anomalies during training. However, the method is designed to function in a self-supervised manner, meaning it generates pseudo-negative samples from normal data to train the model, which suggests that it could be applied to scenarios where anomaly labels are not available. This approach allows the model to differentiate between normal and abnormal behaviors based on the learned representations, potentially making it useful for zero-shot anomaly detection in multivariate time series data. Nonetheless, further validation would be needed to confirm its effectiveness specifically for zero-shot tasks.

The ContrastAD framework presented by B. Li and Müller (2023) is a self-supervised method for time series anomaly detection that leverages contrastive learning with temporal transformations. The key innovation is the use of anomaly-induced transformations to create representations that differentiate between normal and abnormal data. This approach targets both point anomalies and contextual anomalies in high-dimensional time series, which are often missed by other methods. By learning distinct representations for normal and anomalous data in the latent space, ContrastAD improves performance on noisy and complex datasets. However, the method is not trained or validated on entirely new types of anomalies.

Another model called DCdetector presented by Yang, Zhang, Zhou, Wen, and Sun (2023) is a multi-scale dual attention contrastive learning framework designed for time-series anomaly detection. It utilizes a dual attention asymmetric design to create a permutation-invariant representation, guiding the learning process with pure contrastive loss. This approach enhances the model’s ability to discriminate between normal and anomalous data. Extensive experiments demonstrate that DCdetector achieves state-of-the-art performance across multiple benchmark datasets. While the paper focuses on its effectiveness in anomaly detection, it does not explicitly address or test the model’s applicability to zero-shot learning scenarios (Yang et al., 2023).

Pranavan, Sim, Ambikapathi, and Ramasamy (2022) present a novel approach for anomaly detection in multi-variate time series data using Contrastive Predictive Coding (CPC). Their method, named Time-series Representational Learning through Contrastive Predictive Coding (TRL-CPC), aims to effectively capture the temporal dependencies and correlations across multiple variables in time series data. The TRL-CPC framework consists of an encoder, an auto-regressive model, and a non-linear transformation model. These components are jointly optimized to learn the representations of multi-variate time series data by predicting future segments from past segments. The core idea is to maximize the mutual information between the encoded representations of past and future segments, thereby learning robust representations. To detect anomalies, TRL-CPC calculates the prediction error between actual future segments and the predicted segments generated by the CPC model. Anomalies are identified where this prediction error exceeds a certain threshold, enabling unsupervised anomaly detection based on the structure of the data itself. Experimental results show that TRL-CPC outperforms traditional anomaly detection methods on several benchmark datasets, highlighting its effectiveness in capturing complex temporal dependencies and identifying anomalies in multi-variate time series data (Pranavan et al., 2022).

CPC is also used by the method TiCTok presented by Kang and Lee (2023). The model proposes a novel approach to time-series anomaly detection by combining contrastive tokenization with a time-series token encoder. This encoder converts raw time-series data into latent embeddings that capture wide-ranging temporal information. The model employs contrastive learning to produce high-quality representations, which help distinguish between normal and anomalous data. Additionally, TiCTok introduces a new anomaly scoring method based on the contrastive loss used during training. Experimental results indicate that TiCTok performs effectively across multiple benchmark datasets, achieving results that are either superior or comparable to existing state-of-the-art methods. In their paper they do not test the model on Zero Shot Learning scenarios like the previous method.

A paper by ? (?) introduces Multiview Graph Contrastive Learning for detecting anomalies in multivariate time-series data, particularly in IoT systems. The method constructs graph structures to model both temporal context and signal dependencies, while an adaptive data augmentation strategy generates graph

views for contrastive learning. This approach enhances representation quality and improves performance in anomaly detection tasks, outperforming existing methods on multiple real-world datasets

The paper by Y. Sun et al. (2023) introduces TriAD (Tri-domain Anomaly Detector), a self-supervised learning method for time-series anomaly detection. TriAD models features across three domains—temporal, frequency, and residual—without relying on labeled anomalies. Unlike traditional contrastive learning, TriAD uses inter-domain and intra-domain contrastive losses to learn shared attributes among normal data and distinguish them from anomalies. The approach is designed to handle anomalies of varying lengths and shapes, and it achieves significant improvements over state-of-the-art deep learning models in anomaly detection tasks (Y. Sun et al., 2023).

AE "Autoencoders are robust to unclean training datasets" (Abdulaal, 2021, p. 2487).

? (?) are the first to use a Unified Autoencoder (UAE) for time series data, namely the power forecast of wind and solar plants. They contribute to the challenge of predicting the possible outcome of renewable energy in a newly created plant, either wind or solar. To do so a UAE is combined with a Task Embedding Neural Network (TENN) They examine the usability divided in Single-Task, Multi-Task and Zero-Shot Learning. The method was first published in ? (?). It is then extended by convolutional layers instead of the fully connected neural network layers (UCAE-TENN) and also Long Short-Term Memory layers (ULAE-TENN).

Realising few-shot anomaly detection of images is done by ? (?). The method MAEDAY can detect objects newly added to the frames. To achieve this a masked autoencoder is used who recreates the former image but without the anomaly. The difference between the initial and reconstructed images is calculated and the object then visible. This method is useful for its ability to detect anomalies with very few examples, making it a powerful tool in scenarios where labeled data is rare. ? (?) demonstrate the effectiveness of MAEDAY in various applications, showcasing its potential for real-world anomaly detection tasks.

To detect anomalies in healthcare data a variational recurrent autoencoder is used by Pereira and Silveira (2019). The focus is on electrocardiogram (ECG) datasets. Their method tackles the challenge of finding anomalies in unlabelled time series data. They created an unsupervised framework where the model learns to represent the data and detect anomalies without needing labeled examples. The VRAE model works by learning to reconstruct the input sequences. During training, they add noise to the input data, and the model tries to reconstruct the original, uncorrupted data. This helps the model learn more robust representations of the data. To detect anomalies, they cluster these learned representations and use the Wasserstein distance to identify outliers. Their approach was tested on the ECG5000 dataset and showed that it could effectively detect unusual heartbeats, performing better than previous methods that required labeled data.

Another approach using VRAE involves creating synthetic anomalies to improve the detection process. In their method, they use a two-level hierarchical latent space representation. First, they distill feature descriptors of normal data points into more robust representations using autoencoders (AEs). These representations are then refined using a variational autoencoder (VAE) that creates a family of distributions. From these distributions, they select those that lie on the outskirts of the normal data as generators of synthetic anomalies. By generating these synthetic anomalies, they train binary classifiers to distinguish between normal and abnormal data. Their hierarchical structure for feature distillation and fusion helps create robust representations, enabling effective anomaly detection without needing actual anomalies during training. Their method performs well on several benchmarks for anomaly detection (Ramirez Rivera, Khan, Bekkouch, & Sheikh, 2022).

The paper by Zhang et al. (2020) addresses the challenge of detecting unknown cyberattacks by applying zero-shot learning. The proposed method maps the features of known attacks to a semantic space using a sparse autoencoder and restores them to the feature space by minimizing reconstruction errors, effectively creating a mapping between features and semantic attributes. This technique enables the model to detect previously unseen attacks by generalizing from known attack features. The method was tested on the NSL-KDD dataset, achieving an accuracy of 88.3%, outperforming traditional approaches in detecting unknown attacks. The research highlights the feasibility and effectiveness of zero-shot learning for cybersecurity applications.

? (?) point out, that AE-based methods have remaining challenges.

T (Xu, Wu, Wang, & Long, 2022) use a Transformer architecture to detect anomalies on three different datasets. TODO

To overcome the challenge of poorly available time series data sets (Ma et al., 2023), the model family MOMENT tries to learn general patterns on a pile of time series data (Goswami et al., 2024). The pile is a collection of different datasets which they assembled for their pretraining. According to the paper minimal finetuning is needed to perform well on time series tasks like anomaly detection. They published the model and made the usage easily accessible with its own python library. The constructed time serie pile consists of a widespread list of domains including weather measurements, sensor values and power consumption datasets. They also included data not connected with the previous like the tongue and finger movement of humans. The different tasks which the model is evaluated on are forecasting (long and short horizon), classification, anomaly detection and imputation. Except for short-horizon forecasting all tasks are managed well.

Based on the transformer architecture, Large Language Models (LLM) are developed and used increasingly in different applications. Known as chatbots they can help in language specific tasks. Beside that they can be used in anomaly detection and forecasting. J. Su et al. (2024) examine a literature review on how LLMs perform on anomaly detection tasks concerning time series data. LLMs

in anomaly detection are specifically useful when the time series data is in the form of words. This can be the case in log analysis. Logs are generated over time and hold a lot of information which can signify errors and system failures. They conclude that LLMs have potential in detecting anomalies but challenges remain. The occurrence of hallucinations and the need for computational efficiency to name a few (J. Su et al., 2024).

A method detecting anomalies in tabular data is presented by A. Li et al. (2024). In order to perform tasks that LLMs are not directly build for they generate synthetic datasets. Using these datasets LLMs and specifically GPT-4 have comparable performance with transductive learning methods (A. Li et al., 2024, p. 6).

Using Fourier Analysis and the transformer architecture Ye, He, Zhang, Xiao, and Li (2023) detect anomalies in time series data. The encoder of a transformer is used to capture temporal features of time series. They detect frequencies by using Fourier Analysis.

SL Beggel, Kausler, Schiegg, Pfeiffer, and Bischl (2019) address the problem of detecting anomalies in time series data using a novel unsupervised method based on shapelet learning. This approach is particularly useful in scenarios where labeling data is difficult and expensive. Their method learns representative features that describe the shape of time series data from the normal class and simultaneously learns to accurately detect anomalies. The objective function encourages the learning of a feature representation in which normal time series lie within a compact hypersphere, while anomalous observations lie outside the decision boundary. This is achieved through a block-coordinate descent procedure. The advantage of their approach is that it can efficiently detect anomalies in unseen test data without retraining the model, by reusing the learned feature representation. Experimental results on multiple benchmark datasets demonstrate the robustness and reliability of the method in detecting anomalous time series, outperforming competing methods when the training data contains anomalies (Beggel et al., 2019).

In contrast, Alshaer, Garcia-Rodriguez, and Gouy-Pailler (2020) propose a method combining matrix profiles with shapelet learning to handle streaming time series data. The matrix profile efficiently identifies potential anomalies in real-time, and shapelet learning characterizes these anomalies accurately. This approach is particularly suited for environments requiring immediate anomaly detection, such as finance, healthcare, and industrial monitoring (Alshaer et al., 2020).

While both methods utilize shapelet learning, Beggel et al. focus on static datasets and robust feature representation, whereas Alshaer et al. emphasize real-time detection in dynamic, streaming environments.

C ? (?) propose a method for detecting anomalies in multivariate time series by combining clustering and reconstruction techniques. The authors use a sliding window approach to generate subsequences from the multivariate time series,

then apply extended fuzzy clustering to reveal the underlying structure of the subsequences. By reconstructing these subsequences with optimal cluster centers, the method detects anomalies based on how well the reconstructed data fits the original subsequences. A confidence index quantifies the level of detected anomalies. The approach is shown to effectively detect anomalies related to amplitude and shape changes in various application domains, such as healthcare and finance, and is optimized using Particle Swarm Optimization.

5 Application for Zero Shot Anomaly Detection on Multi Variate Time Series Data

In this chapter the found articles are filtered using the exclusion criteria defined in 2.2. By excluding methods that haven't been tested with multiple input variables we answer RQ2. Methods that are not tested in Zero-Shot Learning Scenarios are also excluded which covers RQ4. More precisely we want to know by the defined filter process which of the proposed RL types are best suited for Zero Shot Anomaly Detection in multi-variate time series data. In this chapter a selection of appropriate methods for Time Series Data Anomaly Detection out of 4 is extracted. Here the priors described in 3.1 are used in order to rate the RL types. In order to achieve a successful implementation in the next chapter, the focus is on the opensource availability of the described models. Only models which are available and well documented are chosen for further examination. In 1 the links to public repositories are listed. They are found by citations in the paper.

TODO (Fung, Qiu, Li, & Rudolph, 2024)

6 Proof of Concept

The best fitting strategies are implemented on a small test data set in order to demonstrate how it works.

TODO Include Link to code repo

6.1 Data Set including Anomalies

Which data set to choose for a valid proof of concept. The structure of the chosen data set is described in this chapter.

While NLP and image processing tasks are common and a variety of data sets exists, time series data sets are not available that much (Ma et al., 2023). The transferability between time series datasets is difficult due to the fact that the data between domains is huge (Ma et al., 2023)

In the test data the learning data is separate from the data including anomalies. The important thing about Zero Shot Learning is that a specific anomaly never occurred like this before. In the test data, all chosen representation learning techniques are applied using the same data for learning and afterwards testing

Table 1. Representation learning methodologies and the classification by exclusion criteria. Methods tested with multiple input variables (MV). Methods tested with anomalies which never occurred in the training data (ZSL). Links to repositories containing the models are listed under Implementation.

Method Name	Author	Concept	MV	ZSL	Implementation
INRAD	Jeong et al.	MLP	no	good	
OmniAnomaly	Su et al.	RNN	no		github.com/smallcowbaby/OmniAnomaly
	Kravchik et al.	CNN	no		
	He et al.	CNN	no	difficult	
	Zhang et al.	CL	yes	difficult	
CARLA	Darban et al.	CL	no		
CL-TAD	Ngu et al.	CL	yes		
COCA	Wang et al.	CL	no		
	Lee et al.	CL			
	Chen et al.	CL	no		
TS2Vec	Yue et al.	CL			github.com/yuezhihan/ts2vec
TimeAutoAD	Jiao et al.	CL	no		
ContrastAD	Li et al.	CL	no		
Dcdetector	Yang et al.	CL	no		
TRL-CPC	Pranavan et al.	CL	no		
TiCTok	Kang et al.	CL	no		
MGCLAD	Qin et al.	CL	no		
TriAD	Sun	CL	no		
UCAE-TENN	Nivarthi et al.	AE			
MAEDAY	Schwartz et al.	AE			
	Pereira et al.	AE			
	Ramirez et al.	AE			
MOMENT	Goswami et al.	T	yes	python library	
	Li et al.				
	Beggel et al.	SL			
	Alshaer et al.	SL			
TranAD	Aota et al.				
	Tuli et al.	T			github.com/imperial-quore/TranAD
AnomalyTransformer	Ye et al.	T			
	Xu et al.	T			github.com/thuml/Anomaly-Transformer
RANSynCoders	Abdulaal et al.	AE			github.com/eBay/RANSynCoders
	Li et al.	C			
	Zhang et al.	AE	yes		

the anomaly detection with the same anomalies. According to chapter (Evaluation) the characteristics are evaluated for each RL technique chosen in the previous chapter.

To test the model with anomalies in a consistent data set the Server Machine Dataset (SMD) provided by Y. Su et al. (2019) is used. The SMD is a 5-week-long dataset made up by data from 28 different machines. The anomalies are pre labeled.

how to inject artificial (Darban et al., 2024) /real anomalies. Which real world scenarios do we have? Are there anomalies in SMA data?

6.2 Results

7 Summary

7.1 Discussion

That the methods are not tested on Zero-Shot Scenarios in the paper they are presented doesn't mean they can not perform well on them. Further research and test with the most promising models need to be done in the future.

That is not the case for multivariate data. If a method designed for time series data with a single input variable needs to be used for multiple input variables a redesign is necessary. This can be extensive or rather simple depending on the architecture. To know how much redesign is necessary every method has to be reviewed in further research.

7.2 Future Work

References

- Abdulaal, A. (2021). Practical Approach to Asynchronous Multivariate Time Series Anomaly Detection and Localization.
- Alshaer, M., Garcia-Rodriguez, S., & Gouy-Pailler, C. (2020, November). Detecting Anomalies from Streaming Time Series using Matrix Profile and Shapelets Learning. In *2020 IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 376–383). Baltimore, MD, USA: IEEE. Retrieved 2024-07-30, from <https://ieeexplore.ieee.org/document/9288261/> doi: <https://doi.org/10.1109/ICTAI50040.2020.00066>
- Aota, T., Tong, L. T. T., & Okatani, T. (2023, January). Zero-shot versus Many-shot: Unsupervised Texture Anomaly Detection. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 5553–5561). Waikoloa, HI, USA: IEEE. Retrieved 2024-08-01, from <https://ieeexplore.ieee.org/document/10030870/> doi: <https://doi.org/10.1109/WACV56688.2023.00552>
- Beggel, L., Kausler, B. X., Schiegg, M., Pfeiffer, M., & Bischl, B. (2019, September). Time series anomaly detection based on shapelet learning. *Computational Statistics*, 34(3), 945–976. Retrieved 2024-07-30, from <http://link.springer.com/10.1007/s00180-018-0824-9> doi: <https://doi.org/10.1007/s00180-018-0824-9>
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Chen, K., Feng, M., & Wirjanto, T. S. (2023, April). *Time-series Anomaly Detection via Contextual Discriminative Contrastive Learning*. arXiv. Retrieved 2024-08-27, from <http://arxiv.org/abs/2304.07898> (arXiv:2304.07898 [cs])
- Darban, Z. Z., Webb, G. I., Pan, S., Aggarwal, C. C., & Salehi, M. (2024, April). *CARLA: Self-supervised Contrastive Representation Learning for Time Series Anomaly Detection*. arXiv. Retrieved 2024-07-18, from <http://arxiv.org/abs/2308.09296> (arXiv:2308.09296 [cs])
- Fung, C., Qiu, C., Li, A., & Rudolph, M. (2024, February). *Model Selection of Zero-shot Anomaly Detectors in the Absence of Labeled Validation Data*. arXiv. Retrieved 2024-08-01, from <http://arxiv.org/abs/2310.10461> (arXiv:2310.10461 [cs])
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, Massachusetts: The MIT Press.
- Goswami, M., Szafer, K., Choudhry, A., Cai, Y., Li, S., & Dubrawski, A. (2024, May). *MOMENT: A Family of Open Time-series Foundation Models*. arXiv. Retrieved 2024-07-18, from <http://arxiv.org/abs/2402.03885> (arXiv:2402.03885 [cs])
- Gruhl, C. M. (2022). Novelty Detection for Multivariate Data Streams

- with Probabilistic Models. Retrieved 2024-07-15, from <https://kobra.uni-kassel.de/handle/123456789/13902> (Publisher: Universität Kassel) doi: <https://doi.org/10.17170/KOBRA-202205106160>
- Gärdenfors, P. (2000). *Conceptual spaces: the geometry of thought*. Cambridge, Mass: MIT Press.
- He, Y., & Zhao, J. (2019). Temporal Convolutional Networks for Anomaly Detection in Time Series. *Journal of Physics*.
- Jeong, K.-J., & Shin, Y.-M. (2022, January). *Time-Series Anomaly Detection with Implicit Neural Representation*. arXiv. Retrieved 2024-07-30, from <http://arxiv.org/abs/2201.11950> (arXiv:2201.11950 [cs])
- Jiao, Y., Yang, K., Song, D., & Tao, D. (2022, May). TimeAutoAD: Autonomous Anomaly Detection With Self-Supervised Contrastive Loss for Multivariate Time Series. *IEEE Transactions on Network Science and Engineering*, 9(3), 1604–1619. Retrieved 2024-08-01, from <https://ieeexplore.ieee.org/document/9705079/> doi: <https://doi.org/10.1109/TNSE.2022.3148276>
- Kang, M., & Lee, B. (2023). TiCTok: Time-Series Anomaly Detection With Contrastive Tokenization. *IEEE Access*, 11, 81011–81020. Retrieved 2024-08-27, from <https://ieeexplore.ieee.org/document/10201844/> doi: <https://doi.org/10.1109/ACCESS.2023.3301140>
- Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009, January). Systematic literature reviews in software engineering – A systematic literature review. *Information and Software Technology*, 51(1), 7–15. Retrieved 2024-08-19, from <https://linkinghub.elsevier.com/retrieve/pii/S0950584908001390> doi: <https://doi.org/10.1016/j.infsof.2008.09.009>
- Kravchik, M., & Shabtai, A. (2018, December). *Detecting Cyberattacks in Industrial Control Systems Using Convolutional Neural Networks*. arXiv. Retrieved 2024-08-21, from <http://arxiv.org/abs/1806.08110> (arXiv:1806.08110 [cs])
- Lee, Y., Byun, Y., & Baek, J.-G. (2023, February). Time Series Anomaly Detection Using Contrastive Learning based One-Class Classification. In *2023 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)* (pp. 330–335). Bali, Indonesia: IEEE. Retrieved 2024-08-01, from <https://ieeexplore.ieee.org/document/10067089/> doi: <https://doi.org/10.1109/ICAIIIC57133.2023.10067089>
- Li, A., Zhao, Y., Qiu, C., Kloft, M., Smyth, P., Rudolph, M., & Mandt, S. (2024, June). *Anomaly Detection of Tabular Data Using LLMs*. arXiv. Retrieved 2024-08-01, from <http://arxiv.org/abs/2406.16308> (arXiv:2406.16308 [cs])
- Li, B., & Müller, E. (2023, June). Contrastive Time Series Anomaly Detection by Temporal Transformations. In *2023 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). Gold Coast, Australia: IEEE. Retrieved 2024-08-27, from <https://ieeexplore.ieee.org/document/10191358/> doi: <https://doi.org/10.1109/IJCNN54540>

- .2023.10191358
- Ma, Q., Liu, Z., Zheng, Z., Huang, Z., Zhu, S., Yu, Z., & Kwok, J. T. (2023, May). *A Survey on Time-Series Pre-Trained Models*. arXiv. Retrieved 2024-07-18, from <http://arxiv.org/abs/2305.10716> (arXiv:2305.10716 [cs])
- Ngu, H. C. V., & Lee, K. M. (2023, October). CL-TAD: A Contrastive-Learning-Based Method for Time Series Anomaly Detection. *Applied Sciences*, 13(21), 11938. Retrieved 2024-08-01, from <https://www.mdpi.com/2076-3417/13/21/11938> doi: <https://doi.org/10.3390/app132111938>
- Nielsen, M. (2015). *Neural Networks and Deep Learning*. , 224.
- Palatucci, M., Pomerleau, D., Hinton, G. E., & Mitchell, T. M. (2009). Zero-shot Learning with Semantic Output Codes.
- Pereira, J., & Silveira, M. (2019, February). Learning Representations from Healthcare Time Series Data for Unsupervised Anomaly Detection. In *2019 IEEE International Conference on Big Data and Smart Computing (BigComp)* (pp. 1–7). Kyoto, Japan: IEEE. Retrieved 2024-07-30, from <https://ieeexplore.ieee.org/document/8679157/> doi: <https://doi.org/10.1109/BIGCOMP.2019.8679157>
- Pranavan, T., Sim, T., Ambikapathi, A., & Ramasamy, S. (2022, February). *Contrastive predictive coding for Anomaly Detection in Multi-variate Time Series Data*. arXiv. Retrieved 2024-07-30, from <http://arxiv.org/abs/2202.03639> (arXiv:2202.03639 [cs])
- Ramirez Rivera, A., Khan, A., Bekkouch, I. E. I., & Sheikh, T. S. (2022, January). Anomaly Detection Based on Zero-Shot Outlier Synthesis and Hierarchical Feature Distillation. *IEEE Transactions on Neural Networks and Learning Systems*, 33(1), 281–291. Retrieved 2024-07-30, from <https://ieeexplore.ieee.org/document/9228891/> doi: <https://doi.org/10.1109/TNNLS.2020.3027667>
- Sabokrou, M. (2018). Deep-anomaly_ Fully convolutional neural network for fast anomaly detection in crowded scenes. *Computer Vision and Image Understanding*.
- Socher, R., Ganjoo, M., Sridhar, H., Bastani, O., Manning, C. D., & Ng, A. Y. (2013, March). *Zero-Shot Learning Through Cross-Modal Transfer*. arXiv. Retrieved 2024-07-14, from <http://arxiv.org/abs/1301.3666> (arXiv:1301.3666 [cs])
- Su, J., Jiang, C., Jin, X., Qiao, Y., Xiao, T., Ma, H., ... Lin, J. (2024, February). *Large Language Models for Forecasting and Anomaly Detection: A Systematic Literature Review*. arXiv. Retrieved 2024-08-19, from <http://arxiv.org/abs/2402.10350> (arXiv:2402.10350 [cs])
- Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., & Pei, D. (2019, July). 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* (pp. 2828–2837). Anchorage AK USA: ACM. Retrieved 2024-08-01, from <https://dl.acm.org/doi/10.1145/3292500.3330672> doi: <https://doi.org/10.1145/3292500.3330672>

- doi.org/10.1145/3292500.3330672
- Sun, Q., & Ge, Z. (2021, September). A Survey on Deep Learning for Data-Driven Soft Sensors. *IEEE Transactions on Industrial Informatics*, 17(9), 5853–5866. Retrieved 2024-07-14, from <https://ieeexplore.ieee.org/document/9329169/> doi: <https://doi.org/10.1109/TII.2021.3053128>
- Sun, Y., Pang, G., Ye, G., Chen, T., Hu, X., & Yin, H. (2023, November). *Unraveling the "Anomaly" in Time Series Anomaly Detection: A Self-supervised Tri-domain Solution*. arXiv. Retrieved 2024-08-27, from <http://arxiv.org/abs/2311.11235> (arXiv:2311.11235 [cs])
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is All you Need.
- Wang, R., Liu, C., Mou, X., Gao, K., Guo, X., Liu, P., ... Liu, X. (2023, April). *Deep Contrastive One-Class Time Series Anomaly Detection*. arXiv. Retrieved 2024-08-27, from <http://arxiv.org/abs/2207.01472> (arXiv:2207.01472 [cs])
- Xu, J., Wu, H., Wang, J., & Long, M. (2022, June). *Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy*. arXiv. Retrieved 2024-07-18, from <http://arxiv.org/abs/2110.02642> (arXiv:2110.02642 [cs])
- Yang, Y., Zhang, C., Zhou, T., Wen, Q., & Sun, L. (2023, August). DCdetector: Dual Attention Contrastive Representation Learning for Time Series Anomaly Detection. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* (pp. 3033–3045). Retrieved 2024-08-27, from <http://arxiv.org/abs/2306.10347> (arXiv:2306.10347 [cs]) doi: <https://doi.org/10.1145/3580305.3599295>
- Ye, Y., He, Q., Zhang, P., Xiao, J., & Li, Z. (2023). Multivariate Time Series Anomaly Detection with Fourier Time Series Transformer.
- Yue, Z., Wang, Y., Duan, J., Yang, T., Huang, C., Tong, Y., & Xu, B. (2022, February). *TS2Vec: Towards Universal Representation of Time Series*. arXiv. Retrieved 2024-08-03, from <http://arxiv.org/abs/2106.10466> (arXiv:2106.10466 [cs])
- Zhang, Z., Liu, Q., Qiu, S., Zhou, S., & Zhang, C. (2020). Unknown Attack Detection Based on Zero-Shot Learning. *IEEE Access*, 8, 193981–193991. Retrieved 2024-08-25, from <https://ieeexplore.ieee.org/document/9239385/> doi: <https://doi.org/10.1109/ACCESS.2020.3033494>

8 Originality Statement

I hereby declare that the content of this paper is written on my own and sources from literature are declared as such.

Kassel, 15.09.2024

9 AI Assistance Statement

Artificial intelligence is used for learning the foundations to write this paper. During literature research the GPT "Consensus" helped in researching relevant papers and summarized them for the author. The GPT reassured the correctness of paragraphs.

Kassel, 15.09.2024