ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-6654-0540-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICASSP43922.2022.9746646

DETECTING ANOMALY IN CHEMICAL SENSORS VIA REGULARIZED CONTRASTIVE LEARNING

Diaa Badawi ⋆ *Ishaan Bassi*† *Sule Ozev* † *Ahmet Enis Cetin* ⋆

⋆ Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL
† School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ

ABSTRACT

In this work, we present a method for detecting anomalous chemical sensors using contrastive learning-based framework. In many practical systems, an array of multiple chemical sensors are used. Some of the sensors may malfunction due to sensor drift and chemical poisoning. In standard contrastive learning, the aim is to learn representations that will have maximum agreement among data samples of the same concept while having a minimal agreement with data samples from other concepts. In this work, we adapt standard contrastive learning to learning useful representations for out-of-distribution sample detection. Furthermore, we compare the proposed framework with the cosine similarity measure and a novel similarity measure based on the ℓ_1 norm. Our experimental results show that our approach achieves higher AUC scores (93.6%) than baseline methods (90.1%).

***Index Terms*—** anomaly detection, deep learning, contrastive learning, sensor signal processing, chemical sensors

1. INTRODUCTION

Chemical sensory technology provides cheap and mobile solutions to detecting and identifying different gas analytes. Chemical sensors are widely used in ammonia, methane, and Volatile Organic Compound (VOC) detection, which are known to be carcinogenic and main contributors to the greenhouse effect [1, 2, 3, 4, 5, 6, 7, 8]. One of the main challenges facing chemical sensors technology is the fact that sensor responses vary greatly due to process variations and in-field degradation, which is dubbed as sensor drift. Sensor drift can arise due to internal factors which result in a low-frequency change in a sensor response, and due to external reasons such as changes in humidity and temperature. This makes detecting anomalous behavior in the sensory system crucial for reliable gas identification and concentration estimation [9, 10, 11, 12, 13, 14]. In this work, we focus on detecting anomalies in sensor behavior due to external factors in a system consisting of multiple sensors. In [12], the authors propose an online method to detect anomalous changes in wireless sensor measurements by fitting piecewise linear models for the time-series data, and comparing them with reference signals. If the absolute differences are larger than a threshold, they declare anomaly. However, with large sensor-to-sensor variation, such consensus methods may fail to identify the failing sensors and introduce unnecessary uncertainty into the decision-making process.

This work is being supported in part by NSF grants 1739396 (UIC) and 1739451 (ASU). Badawi's work is also partially supported by NSF grant 1934915.

Deep learning architectures such as recurrent neural networks (RNN) and temporal convolution neural networks (TCNN) have been quite popular in learning tasks involving time-series data and anomaly detection [14]. In [15], the authors propose using deep autoencoders to detect sudden changes in a sensor time series in wireless sensor networks. One type of learning, namely contrastive learning, has proven effective in learning pretext tasks that are useful when there is not enough labelled data [16, 17, 18]. In standard contrastive learning, the objective is to learn representations that will have maximum agreement among data samples of the same concept, while having a minimal agreement with data samples from other concepts. Many frameworks have been recently developed such as unsupervised non-parametric Instance discrimination (InstDisc) [16], Momentum Contrast (MoCo) [17], and Simple Contrastive Learning (SimCLR) [18]. In InstDisc, the authors try to maximize the contrast between individual instances across the entire data set. In SimCLR, the authors sample a subset of data points and duplicate each data point by applying an augmentation transform. The objective therein is to maximize the similarity between the data points of each pair against the remaining data points. Unlike previous methods, SimCLR is simple to implement as it avoids the usage of memory banks used in other paradigms. Furthermore, SimCLR achieves state-of-the-art results on ImageNet with a linear classifier.

Motivated by these recent advances, we aim to leverage a deep contrastive framework for detecting an outlier sensor in a sensor array system based on the temporal responses of the sensors. Recent work that applies contrastive learning to detect outliers includes Novelty Detection via Contrastive Learning [19].

In this work, we have actual sensor measurements rather than relying on augmentation transforms to learn similarity/dissimilarity. We modify the standard multi-view contrastive loss criterion to encourage learning representations that are similar among in-distribution samples, (i.e. good sensor time series) while at the same time learning contrasted representations for outlier time series (i.e. degraded sensor time series). In contrastive learning, the cost function uses the cosine similarity measure. We also study an ℓ_1 norm-based similarity measure as a part of contrastive learning for anomalous sensors.

The organization of this paper is as follows: In Sec. 2, we provide the mathematical description of our loss function and our ℓ_1-based similarity metric. In Sec. 3, we provide details on our dataset and our experimental settings and results. Finally, in Sec. 4 we discuss our findings and provide a conclusion of this work.

2. ANOMALY DETECTION NETWORK

In standard multi-view contrastive learning, a random minibatch of size N is sampled, and each sample is augmented by a set of augmentation transforms. This yields a total of $2N$ sample bag. Let $g(z, w)$ be a similarity score (e.g., the cosine similarity measure) between the representation codes z and w. Let (z_i, z_j) be a pair of a samples, where z_i is an image from the minibatch, and z_j represents its augmented version. In [18], the following loss function is minimized:

$$\mathcal{L}_{i,j} := -\log \frac{\exp\big(g(z_i, z_j)/\tau\big)}{\sum_{k=1}^{2N} \mathbb{I}_{i\neq k} \exp\big(g(z_i, z_k)/\tau\big)}, \tag{1}$$

where $\mathbb{I}$ is an indicator function, and $\tau > 0$ is temp hyper-parameter controlling how strong the constrast should be.

2.1. Outlier-Modified Contrastive Loss

The loss function in Eq. 1 is not suitable for anomalous sensor detection problem. This is because the loss in Eq. 1 tries to maximize the similarity between one data instance and its corresponding augmented version one at a time. This is in contrast to our objective to maximize similarity among all in-distribution samples, while maximizing the dissimilarity between the in-distribution and the outlier samples.

Given a set of in-lier sensor measurements $\mathcal{X} = \{x_1, x_2, \dots x_n\}$ and an outlier $\{x^{n+1}\}$. Let $z = f(x)$ be the corresponding features extracted by a neural network. Define p_{ij} for $i \in \{1, 2, ..., n\}$ and $j \in \{1, 2, ..., n+1\}$ as follows

$$p_{ij} = \frac{\exp\big(g(z_i, z_j)/\tau\big)}{\sum_{k=1}^{N} \exp\big(g(z_i, z_k)/\tau\big) + \exp(g(z_i, z^{n+1})/\tau)} \tag{2}$$

Our goal is to have $p_{i,n+1} = 0$ while maximizing the entropy of the in-liers, i.e. minimizing the following objective

$$\mathcal{L}_i = \sum_{l=1,\ l\neq i}^{N} p_{il} \log p_{il} \tag{3}$$

where $p_{i,n+1}$ represents softmax score between an inlier x_i and the outlier x^{n+1}. In our framework, we minimize the unconstrained modification loss criterion defined as follow

$$\mathcal{L}_i = \sum_{l=1, l\neq i}^{N} p_{il} \log p_{il} - \alpha \log(1 - p_{i,n+1}), \tag{4}$$

where $\alpha > 0$. The first term measures the disparity between sample x_i and all other in-lier instances in the minibatch, while the second term measures the repulsion between sample x_i and the outlier sample x^{n+1}. If $p_{i,n+1}$ approaches to 1 the second term grows to infinity. On the other, hand, when $p_{i,n+1}$ approaches zero, as desired, the second term vanishes in Eq. 4. By controlling the Lagrange multiplier α we can put more emphasis on learning invariance among the in-lier samples, in the case of small alpha, while for a large choice, we put more emphasis on differentiating the sample x_i from the outlier point x^{n+1}. The overall minibatch loss is given by:

$$\mathcal{L} := \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}_i \tag{5}$$

The above loss function defined in (4) can be easily modified to accommodate multiple outliers. Let $\{x^{n+1}, x^{n+2}, ..., x^{n+k}\}$ be a set containing k outliers, then the corresponding loss function becomes:

$$\mathcal{L}_i = \sum_{l=1,\ l\neq i}^{N} p_{il} \log p_{il} - \alpha \sum_{j=n+1}^{n+k} \log(1 - p_{i,j}), \tag{6}$$

2.2. Kernel-Based Cosine Similarity Metric

The cosine similarity metric is frequently used to measure the similarity between representations in contrastive learning. The correlation coefficient or the cosine similarity is bounded between -1 and +1. Motivated by this, and by our early work on devising novel dot-product like operations [20], we define the following operator between two vectors $\mathbf{z}$ and $\mathbf{w} \in \mathbb{R}^D$

$$\mathbf{z} \oplus \mathbf{w} := \sum_{i=1}^{D} \text{sgn}(z_i.w_i) \min(|z_i|, |w_i|), \tag{7}$$

to which we refer as *Min-operator*. It also defines a Mercer-type kernel [21]. One can see that the operator is symmetric, continuous and $\mathbf{z} \oplus \mathbf{0} = \mathbf{0} \oplus \mathbf{z} = 0$. Furthermore, $\mathbf{z} \oplus \mathbf{z} = |\mathbf{z}|$, i.e., the operator mimics the regular dot product by "inducing" the ℓ_1-norm instead of the ℓ_2-norm. Based on the above-mentioned properties, we now define a quasi-cosine metric as follows

$$g_{\oplus}(\mathbf{z}, \mathbf{w}) = \frac{\mathbf{z} \oplus \mathbf{w}}{\min(|\mathbf{z}|, |\mathbf{w}|)}, \tag{8}$$

which is bounded by -1 and +1. Furthermore, for any vector $\mathbf{z} = c\mathbf{w}$, $g_{\oplus}(z, w) = \text{sgn}(c)$ similar to the regular cosine similarity measure.

2.3. Inference Phase

During inference, we are presented with a patch of N time series collected by the N sensors in our multi-sensor system. We extract features $z_i = f(x_i)$, where x_i is the i-th time series and $f(.)$ is our neural network. Once we have extract a feature z_i, we find its "distance" scores from the remaining features z_j for $j \neq i$. The notion of distance score is closely related to the similarity metric. We define it to be:

$$d(z_i, z_j) := \frac{1 - g(z_i, z_j)}{2} \tag{9}$$

where $g(.,.)$ is the cosine similarity metric or the metric defined in Eq. 8. As one can see, if the similarity score is close to 1, $d(.,.)$ will be close to zero, and if the similarity score is close to -1, $d(.,.)$ will be close to one. Given all distance scores between feature vector z_i and the remaining features, we now define our anomaly score $S(z_i)$ as follows

$$S(z_i) := \text{Median}\{d(z_i, z_j) : j \neq i\} \tag{10}$$

Note that for a patch of N time series, we need to calculate $\frac{N(N-1)}{2}$ distance scores. If the feature-extracting model has been trained properly, one should be able to find a threshold $T \in (0, 1)$ such that if $s(z) = s\big(f(x)\big) \geq T$, the corresponding sample x should be declared out-of-distribution.

2.4. Feature Extraction Deep Network

We train a temporal convolution neural network with 4 convolutional blocks, each of which has a 3 convolutional layers. The first and the last layers implement causal dilated convolution, while the middle layer implements 1×1 convolution. There are skip (residual) connections between these blocks. The dilation rates are $\{1, 2, 4, 8\}$.

Input	Operator	Filter size	Channels	dilation rate
384×1	Conv1d	3	16	1
384×16	Conv1d	1	32	1
384×32	Conv1d	3	16	1
384×16	Conv1d	3	16	2
384×16	Conv1d	1	32	1
384×32	Conv1d	3	16	2
384×16	Conv1d	3	16	4
384×16	Conv1d	1	32	1
384×32	Conv1d	3	16	4
384×16	Conv1d	3	16	8
384×16	Conv1d	1	32	1
384×32	Conv1d	3	16	8
384×16	Conv2d	1	1	1

Table 1. Temporal convolutional neural network architecture. Each block (except the final layer), consist of 3 layers. There are residual (skip) connections between consecutive blocks.

Finally we have a 1×1 convolutional layer and the output is a 1-dimensional representation z that has the same length as the input x. The details of our model are summarized in Table 1.

3. EXPERIMENTS AND RESULTS

The commercial sensors used for data collection were MQ137 which are Tin oxide(SnO2) based sensors [22]. SnO2 when heated and exposed to air reacts with the oxygen present in air and for a layer of negative ion on the surface and reduce the surface conductivity [22]. However, when gases such as Ammonia come in contact with the surface, they combine with the oxide ion layer on the top and release electrons for conduction and hence increasing the conductivity of the surface. This change in surface resistance can be measured in the form of voltage.

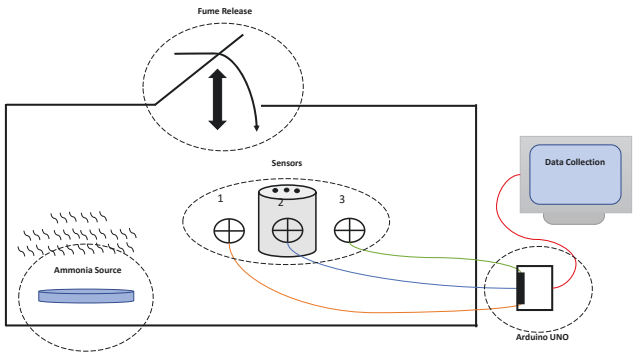

Fig. 1. Illustration of our experimental setup.

Model	AUC score
Deep Contrastive Model (Cosine Similarity)	93,6%
Contrastive Model (Min-operator Similarity)	88,1%
Cosine Similarity (No Feature Extraction)	90,1%
Min-operator Similarity (No Feature Extraction)	89,4%
Absolute Difference [12] (No Feature Extraction)	77,6%

Table 2. Area under curvature (AUC) for various methods. The deep contrastive learning framework provides the best AUC.

3.1. Experiment Setup and Data acquisition

For the experiment, three MQ137 sensors were used. The data was recorded using an Arduino UNO board. The experimental set-up is illustrated in Fig. 1. The three sensors are placed in an airtight chamber. The pre-heating duration of the sensors is 48 hours, during which time no ammonia is introduced into the chamber. A cylindrical ammonia source (commercial low concentrations ammonia cleaner liquid) is placed in the chamber. The lid of the ammonia cylinder is removed to release ammonia gas into the chamber slowly. The ammonia then starts to leak up into the closed environment, gradually building up in the air. Since the chamber is small (¡10L in volume), the concentration of ammonia homogenizes quickly. The sensor responses change accordingly. Based on the proximity of the sensors to one another, they should read the same ammonia concentration. One of the sensors is obstructed with a cylindrical cover with multiple holes. The covering of one of the sensors causes it to react more slowly to the ammonia build-up and release. Therefore, by inducing discrepancy in the sensor response with respect to the other unblocked sensors, we created outlier sensor measurements. The obstruction level of the outlier sensor is modified from experiment to experiment to avoid overfitting to one condition. The chamber lid is opened at random intervals and at random duration to generate a more realistic environment with varying levels of ammonia concentration. We repeated opening and closing the lid multiple times in order to create different rise and fall response.

3.2. Anomaly Detection Example

In the training phase, we segment our time-series into segments of 384 units. We then pre-process each segment by subtracting its mean and dividing by its ℓ_2 norm. This is to eliminate global biases among different sensors. We train a temporal convolutional neural network with the architecture described in Sec. 2.4. We set α in Eq. 4 and τ in Eq. 2 to 0.5 and 1, respectively.

We used measurements corresponding to five experiment sessions for training. Each experiment has three time series belonging to the three sensors, one of which is an outlier (or anomaly). We

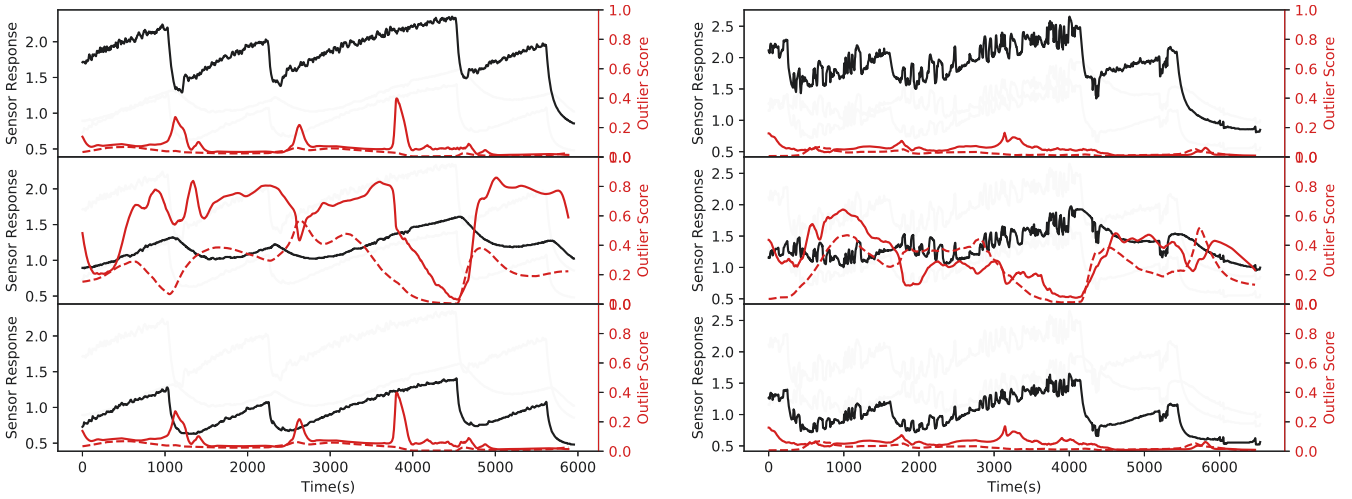

Fig. 2. Outlier score results for the three sensors in two experiments used in testing. The second sensor (second rows) is the poisoned sensor. The learned-representations outlier score is in red, while the dashed red lines correspond to the outlier score with the cosine similarity metric applied directly to the input (no learning). Notice in the case of the second sensor, almost all the time the deep outlier score is significantly higher than in the baseline case. In the right experiment, both scores decrease at around time 4000 seconds. This is because the anomaly experienced at the previous discharge is no longer present in the 384-second-long segments. We consider the output then to be in-lier.

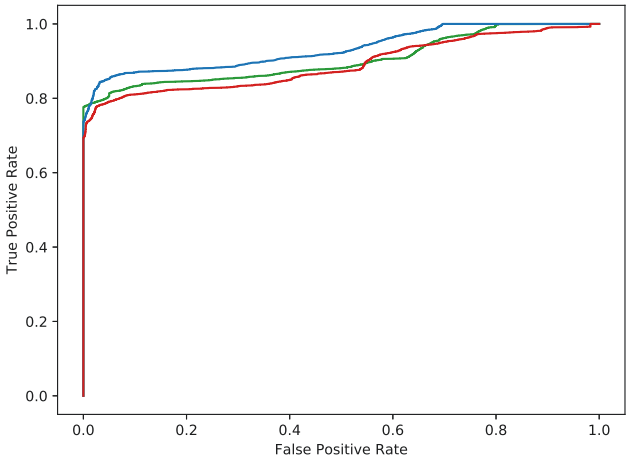

Fig. 3. Receiver operating characteristic curve (ROC) for the contrastive-learning model (blue), shallow cosine-similarity-based model (green), and the shallow min-operator-based similarity (red).

extracted a total of 6000 384-unit-long segments from these time series. We reserved 20% for the data for validation. We used a dataset of four experiments as our test data. During the test phase, we apply three sliding windows of size 384 on each time series coming from the three sensors. We pre-process each segment as in training, feed them into our neural network to extract features. We then calculate the anomaly score as described in Sec. 2.3. For comparison, we applied the cosine similarity, and the min-operator similarity metric defined in Eq. 8 directly on these segments without applying the temporal neural network. Furthermore, we also applied the absolute difference metric $d(z_i, z_j) := \frac{1}{T}\sum_{t=1}^{T} |z_i[t] - z_j[t]|$ used in [12] for anomaly detection. We summarize our accuracy on the test data set in Tab. 2. The best result is obtained by the contratively learned TCNN model as shown in Tab. 2. We plot the Receiver Operating Characteristic (ROC) curve in Fig. 3. By examining the ROC curves, the contrastive-learning-based model can achieve a true positive rate (TPR) of 85% at a false positive rate (FPR) of 2.5%. On the other hand, the other models (the cosine-similarity and min-operator-similarity modes) cannot achieve a TPR higher than 78% at the same FPR.

In this case, the standard cosine similarity based contrastively learned TCNN produces superior results, as opposed to the ℓ_1-norm based methods defined in Eq. (8), we believe in different scenarios, in which the data is contaminated by impulsive noise, the min op-related approach could provide more robust results as in [21].

4. CONCLUSION

In this work, we presented a framework for detecting anomalous sensor(s) in a chemical sensory system using a contrastive learning approach. In this approach, we adapt the standard multi-view contrastive learning loss function such that the model learns to maximize similarity among in-distribution samples (good sensors' readouts) while at the same time maximize dissimilarity between the in-distribution and out-of-distribution samples. We gathered data from three commercial Tin Oxide (SnO2) sensors by exposing them to Ammonia in an environment-controlled experiment. We train a temporal CNN on 4 sets of measurements, and test it on 6 other sets. Our results show that we can identify the anomalous sensor among the three sensors with an AUC score of 93.6%, compared to 90.1% in the baseline case.

5. REFERENCES

[1] Alexander Vergara, Shankar Vembu, Tuba Ayhan, Margaret A Ryan, Margie L Homer, and Ramón Huerta. Chemical gas sensor drift compensation using classifier ensembles. *Sensors and Actuators B: Chemical*, 166:320–329, 2012.

[2] Diaa Badawi, Sule Ozev, Jennifer Blain Christen, Chengmo Yang, Alex Orailoglu, and A Enis Çetin. Detecting gas vapor leaks through uncalibrated sensor based cps. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8296–8300. IEEE, 2019.

[3] Diaa Badawi, Hongyi Pan, Sinan Cem Cetin, and A Enis Çetin. Computationally efficient spatio-temporal dynamic texture recognition for volatile organic compound (voc) leakage detection in industrial plants. *IEEE Journal of Selected Topics in Signal Processing*, 14(4):676–687, 2020.

[4] Diaa Badawi, Agamyrat Agambayev, Sule Ozev, and A Enis Cetin. Real-time low-cost drift compensation for chemical sensors using a deep neural network with hadamard transform and additive layers. *IEEE Sensors Journal*, 2021.

[5] United States. Environmental Protection Agency. Enforcement and Compliance. *Inspection Manual: Federal Equipment Leak Regulations for the Chemical Manufacturing Industry*. U.S. Environmental Protection Agency, Enforcement and Compliance Assurance, 1998.

[6] National Institute of Standards and Technology. Toluene. last accessed: 2019-06-30.

[7] United States Environmental Protection Agency. Leak detection and repair a best practices guide. Accessed: 2019-06-30, url = https://www.epa.gov/sites/production/files/2014-02/documents/ldarguide.pdf, urldate = 2016-12-11.

[8] Fatih Erden, E Birey Soyer, B Ugur Toreyin, and A Enis Cetin. Voc gas leak detection using pyro-electric infrared sensors. In *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1682–1685. IEEE, 2010.

[9] DY Chen and Pak Kwong Chan. An intelligent isfet sensory system with temperature and drift compensation for long-term monitoring. *IEEE Sensors Journal*, 8(12):1948–1959, 2008.

[10] Kow-Ming Chang, Chih-Tien Chang, Kuo-Yi Chao, and Chia-Hung Lin. A novel ph-dependent drift improvement method for zirconium dioxide gated ph-ion sensitive field effect transistors. *Sensors*, 10(5):4643–4654, 2010.

[11] Alisa Rudnitskaya. Calibration update and drift correction for electronic noses and tongues. *Frontiers in chemistry*, 6:433, 2018.

[12] Yuan Yao, Abhishek Sharma, Leana Golubchik, and Ramesh Govindan. Online anomaly detection for sensor systems: A simple and efficient approach. *Performance Evaluation*, 67(11):1059–1075, 2010.

[13] Matt Calder, Robert A Morris, and Francesco Peri. Machine reasoning about anomalous sensor data. *Ecological Informatics*, 5(1):9–18, 2010.

[14] Laura Erhan, M Ndubuaku, Mario Di Mauro, Wei Song, Min Chen, Giancarlo Fortino, Ovidiu Bagdasar, and Antonio Liotta. Smart anomaly detection in sensor systems: A multi-perspective review. *Information Fusion*, 2020.

[15] Tie Luo and Sai G Nagarajan. Distributed anomaly detection using autoencoder neural networks in wsn for iot. In *2018 ieee international conference on communications (icc)*, pages 1–6. IEEE, 2018.

[16] Zhirong Wu, Yuanjun Xiong, Stella X Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3733–3742, 2018.

[17] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9729–9738, 2020.

[18] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[19] Jihoon Tack, Sangwoo Mo, Jongheon Jeong, and Jinwoo Shin. Csi: Novelty detection via contrastive learning on distributionally shifted instances. *arXiv preprint arXiv:2007.08176*, 2020.

[20] Arman Afrasiyabi, Diaa Badawi, Bariş Nasir, Ozan Yıldız, Fatoş T Yarman-Vural, and A Enis Çetin. Non-euclidean vector product for neural networks. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6862–6866. IEEE, 2018.

[21] Hongyi Pan, Diaa Badawi, Erdem Koyuncu, and A Enis Cetin. Robust principal component analysis using a novel kernel related with the l1-norm. *arXiv preprint arXiv:2105.11634, presented in EUSIPCO 2021*, 2021.

[22] Joseph Watson, Kousuke Ihokura, and Gary SV Coles. The tin dioxide gas sensor. *Measurement Science and Technology*, 4(7):711, 1993.