Cross-aligned and Gumbel-refactored Autoencoders for Multi-view Anomaly Detection

Shaoshen Wang, Yanbin Liu, Ling Chen, and Chengqi Zhang
University of Technology Sydney, Sydney, Australia
{shaoshenwang, csyanbin}@gmail.com, {Ling.Chen, Chengqi.Zhang}@uts.edu.au

***Abstract*—Multi-view anomaly detection (AD) is a challenging task due to the complicated data distributions across different views. Specifically, there exist two types of anomalies in multi-view distributions: *attribute anomaly* that exhibits consistent anomalous pattern in each view and *class anomaly* that exhibits inconsistent traits (e.g., semantic label) across multiple views. Existing methods detect these anomalies in an unsupervised manner with the clustering assumption: normal data share consistent clustering structure across views while anomalous data exhibits inconsistent clusters across views. However, these methods would fail for complex multi-view data distributions where there is no obvious clusters. Moreover, existing models suffer from robustness since they are undermined by anomalies during training time. To get rid of the clustering assumption, we propose a Cross-aligned and Gumbel-refactored AutoEncoders (CGAEs) model to effectively detect two types of multi-view anomalies. In CGAEs, we devise a cross-reconstruction module to detect class anomaly by recovering one view from another view. Class anomalies would lead to high cross-reconstruction loss since they do not have the correct information in one view to generate another. We further design a view-alignment module to detect attribute anomaly by the alignment distance among multiple views in the latent space. Attribute anomalies possess large distances since they are less aligned due to fewer anomalous training instances. To handle the robustness issue, we propose a Gumbel-refactored reconstruction loss to replace the mean square error (MSE) in original autoencoders. The cross entropy loss is calculated between the discreterized input and Gumbel-sampled output, thus disregarding the irrelevant details to achieve model robustness. Experimental results validate the superiority of the proposed CGAEs model on both the benchmark datasets and real world datasets. Code will be publicly available.**

***Index Terms*—Anomaly Detection, AutoEncoder, Gumbel**

I. Introduction

Anomaly Detection (AD) [1] aims to identify data points that possess significantly different characteristics from the normal data. While conventional anomaly detection focuses on data from a single view, data in practical applications are often collected from multiple views. For example, social media data usually contain texts, images, and user behaviours that provide complementary information for the same item from different perspectives. Therefore, multi-view anomaly detection emerges as a crucial research problem that finds many real-world applications such as malicious insider detection [2], purchase behavior analysis [3] and disparity management [4].

Multi-view anomaly detection can be categorized into two types [5]. 1) *Attribute anomaly* exhibits consistent abnormal behaviours in each individual view. For example, the yellow triangles shown in Figure 1 are distant from the normal data in each view. 2) *Class anomaly* exhibits inconsistent characteristics across different views. For example, the red squares in Figure 1 behave normally in each view, but show inconsistent patterns if multiple views are considered collectively.

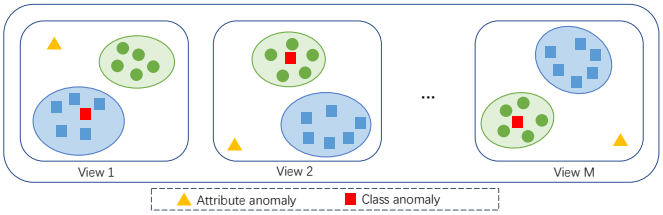


Fig. 1. Illustration of two types of anomalies. *Attribute anomaly* (yellow triangles) exhibits consistent anomalous patterns in each view. *Class anomaly* (red squares) exhibits inconsistent traits across multiple views.

A number of methods have been proposed to detect these multi-view anomalies in an unsupervised manner [2], [3], [5]–[12] (i.e., unsupervised multi-view anomaly detection). Despite the progress, there remain two major issues in these methods. First, most existing approaches rely on the clustering assumption: normal instances share consistent clustering structures across various views while anomalous instances exhibits inconsistent congregating behaviors across different views. However, in practical applications, data instances from multiple views are often of high-dimension and follow complicated distributions, which show no obvious clusters. Consequently, clustering-based approaches [2], [3] may not prevail in such cases. Second, existing methods suffer from the robustness issue since they are undermined by anomalies during model learning. Unsupervised anomaly detection learns the model from the entire dataset including both normal and abnormal instances without label information. Therefore, it can be affected by anomalies because the model tries to fit both anomalous instances and normal ones.

To circumvent the clustering assumption, we adopt an autoencoder as the basic building block in our framework. As a powerful unsupervised learning method, autoencoder can extract and preserve intrinsic information in its discriminative latent codes, which is a favorable option for unsupervised anomaly detection. Specifically, we propose a **Cross-aligned and Gumbel-refactored Autoencoders (CGAEs)** model to

effectively detect the two types of multi-view anomalies. In CGAEs, we first design a *cross-reconstruction module* to detect class anomaly by generating the data of one view from the latent code of another view. Class anomalies exhibit high cross-reconstruction loss since they do not have the correct information in one view to generate another. We then devise a *view-alignment module* to detect attribute anomaly by minimizing the alignment distance of multiple views in the latent space. Attribute anomalies possess large distances since they are less aligned due to the existence of fewer anomalous training instances than normal ones.

To address the model robustness issue, we propose a Gumbel-refactored reconstruction loss to replace the mean square error (MSE) in original autoencoders. Specifically, we reconstruct a discrete version of the input rather than the continuous one to avoid overfitting the irrelevant details. The cross-entropy loss is calculated between the discreterized input and the Gumbel samples from the decoder output. With the discretization and Gumbel-refactored reconstruction, the robustness of autoencoders can be improved by disregarding the unnecessary noises while preserving the indispensable information for anomaly detection.

In the experiments, we verify the effectiveness of each proposed components and compare with various state-of-the-art methods. Our contributions can be summarized as follows:

- We propose a novel **C**ross-aligned and **G**umbel-refactored **A**uto**E**ncoder**s** (CGAEs) model for multi-view anomaly detection. In CGAEs, the cross-reconstruction module and view-alignment module are well-devised to effectively detect the attribute anomaly and class anomaly, respectively.
- We propose a Gumbel-refactored reconstruction loss to replace the original mean square error (MSE) in autoencoders, which endues robustness to the proposed CGAEs for effective anomaly detection.
- We conduct extensive experiments on popular anomaly detection datasets as well as real world data to verify the effectiveness of the proposed method. Experimental results demonstrate the superior performance of CGAEs over the state-of-the-art.

II. Related Work

A. Multi-view Anomaly Detection

Anomaly detection (a.k.a. outlier detection) is an important research topic in machine learning and data mining [1], [13]–[15]. Since anomalies in data usually provide critical information, it has been widely used in various applications, such as health care [16], fraud detection [17], and malicious detection [2]. However, most anomaly detection approaches are designed for single-view data.

Multi-view anomaly detection aims to detect anomalous instances by considering the data distributions across multiple views. It is a practical and challenging research task that has attracted an increasing amount of attention. Earlier works [18], [19] focus on detecting outliers that exhibit abnormal behaviours in each individual view. Subsequent studies shift the focus to identify inconsistent cross-view cluster membership [2], [3], [20]. For example, HOAD [3] first clusters objects in each view by spectral clustering [21] and then identifies anomalies as the samples belonging to different clusters in different views. Works in [2], [20] aim at detecting outliers by exploring inconsistency of clustering across multiple views with consensus clustering and affinity propagation [22], respectively. By design, these methods are constrained to detect class anomaly only. Similar to the clustering assumption, recent works [5] propose a nearest neighbor based outlier measurement criterion, which assumes that multiple views of a normal instance have similar neighborhood structures while anomalies do not. Different from aforementioned approaches, our proposed method does not rely on strong clustering or nearest-neighbor assumptions, thereby being more flexible to handle the complex multi-view data distributions.

Most existing multi-view anomaly detection algorithms deal with the problem in an unsupervised manner (a.k.a. unsupervised multi-view anomaly detection), considering it is either costly or infeasible to obtain sufficient labelled data in real world applications. Works in [6], [23] use low-rank matrix recovery to exploit the representation of both normal and anomalous data and identifies anomalies by measuring the consistency of representation coefficients in each view. MODDIS [9] integrates all multi-view data into a latent intact space to preserve sufficient irregularities of samples. Since the entire dataset including both normal data and anomalies are used during model learning, these methods suffer from the robustness issue as the learned model is subject to the influence of anomalies. In this paper, we propose a Gumbel-refactored reconstruction loss to replace the original mean square error (MSE) in autoencoders. The reconstruction between the discreterized input and Gumbel-sampled output prevents our model from capturing noisy and irrelevant details of the data, leading to robust anomaly detection.

B. Autoencoders for Anomaly Detection

Autoencoders have been widely-used in anomaly detection [24]–[27]. The general idea is to train an autoencoder on the entire dataset (both normal and anomalous) and utilize the reconstruction error as the detection criterion. As there are fewer anomalous data for training, the reconstruction errors of anomalies are higher than that of normal ones, which can be utilized to separate the anomalies from normal data. DAGMM [24] improves this idea with Gaussian Mixture Model (GMM) and jointly optimizes the parameters of the deep autoencoder and the mixture model. MemAE [25] proposes a memory-augmented autoencoder to improve the performance of unsupervised anomaly detection. However, all these methods focus only on single-view anomaly detection. On the contrary, our CGAEs model introduces the cross-reconstruction loss and view-aligment loss, which are well-devised for detecting two types of anomalies in multi-view data. In the context of multi-view anomaly detection, very few autoencoder-based methods have been proposed. For example, 3D-STAE [28] proposes two variants of the spatio-temporal

autoencoder based on handcrafted and deep features for video anomaly detection. While 3D-STAE is specifically designed for the video anomaly detection task, our CGAEs model is a general multi-view anomaly detection method that can be applied to multiple machine learning tasks (e.g., tasks in UCI machine learning repository) and real-world tasks (e.g., document anomaly in Newsgroup datasets [29]).

III. Methodology

A. Problem Setting

Suppose that we are given a dataset D including N instances $\{X_1, ..., X_N\}$ with M views. $X_i = (X_i^{(1)}, \ldots, X_i^{(M)})$ is a set of multi-view observation vectors of the i-th instance, and $X_i^{(m)} \in \mathbb{R}^{d_m}$ is the observation vector of the m-th view with d_m being its dimension. The goal of unsupervised multi-view AD is to find anomalous instances that have inconsistent characteristics across multiple views. In this paper, we focus on detecting two types of anomalies: *attribute anomaly* and *class anomaly*. Without loss of generality, we first introduce the proposed model to deal with two views (i.e., $M = 2$) by the proposed Cross-aligned Autoencoders and Gumbel-refactored Reconstruction. Then, we describe an efficient strategy to extend our model to handle more than two views.

B. Cross-aligned Autoencoders

We start by analysing the behaviours of the normal and anomalous instances. Given the data of two views, we denote a normal instance as $X_i = [X_i^{(1)}, X_i^{(2)}]$ and an anomalous instance as $X_j = [X_j^{(1)}, X_j^{(2)}]$. For multi-view anomaly detection, we explore the following two natural assumptions.

First, **the correlation among multiple views of a normal instance is stronger than that of an anomalous instance.** For example, considering an image and its description text as two views, the correlation between the correct image-text views should be much stronger than the correlations between incorrect image-text views. Specifically, we model the cross-view correlations using the conditional distribution of different views, i.e., $p(X^{(2)}|X^{(1)})$ and $p(X^{(1)}|X^{(2)})$. Then, this assumption can be formulated as

$$p(X_i^{(2)}|X_i^{(1)}) + p(X_i^{(1)}|X_i^{(2)}) > p(X_j^{(2)}|X_j^{(1)}) + p(X_j^{(1)}|X_j^{(2)}).$$

The conditional distributions can be modeled through a proper cross-view generation process.

Second, **in the latent space, the distance among multiple views of a normal instance should be smaller than that of an anomalous instance.** Formally, let $z^{(1)}$ and $z^{(2)}$ be the latent embeddings of $X^{(1)}$ and $X^{(2)}$, then

$$\|z_i^{(1)} - z_i^{(2)}\| < \|z_j^{(1)} - z_j^{(2)}\|,$$

where $\|\cdot\|$ is a distance metric, e.g., Euclidean distance. This assumption means that multi-view features of a normal instance are more consistent to be projected into nearby positions in the latent space.

To incorporate the above assumptions into the model design, we propose the Cross-Aligned AutoEncoders (CAAE), which consists of a *cross-reconstruction module* (motivated from assumption 1) and a *view-alignment module* (motivated from assumption 2), as shown in in Figure 2. In the cross-reconstruction module, the conditional distribution is naturally modeled as the cross-view reconstruction probability in autoencoders. Concretely, to calculate $p(X^{(2)}|X^{(1)})$, $X^{(1)}$ is input to an encoder E_{1c} to get the latent code $z_c^{(1)}$, followed by a decoder D_{1c} to obtain the reconstruction $X'^{(2)}$.

$$z_c^{(1)} = E_{1c}(X^{(1)}),\ X_c'^{(2)} = D_{1c}(z_c^{(1)}), \tag{1}$$

$$p(X^{(2)}|X^{(1)}) = p(X^{(2)}|X_c'^{(2)}). \tag{2}$$

Similarly, we can calculate the symmetric conditional probability $p(X^{(1)}|X^{(2)}) = p(X^{(1)}|X_c'^{(1)})$ with an encoder E_{2c} and a decoder D_{2c}. In order to train these encoders and decoders, we utilize the maximum log likelihood criterion to introduce the cross-reconstruction loss as

$$\mathcal{L}_{\mathrm{C}} = -[\log p(X^{(1)}|D_{2c}(z_c^{(2)})) + \log p(X^{(2)}|D_{1c}(z_c^{(1)}))]. \tag{3}$$

In the view-alignment module, we introduce another pairs of encoders and decoders to reconstruct each view as

$$z_a^{(1)} = E_{1a}(X^{(1)}),\ X_a'^{(1)} = D_{1a}(z_a^{(1)}), \tag{4}$$

$$z_a^{(2)} = E_{2a}(X^{(2)}),\ X_a'^{(2)} = D_{2a}(z_a^{(2)}). \tag{5}$$

To incorporate the second assumption, we introduce the following view-alignment loss

$$\mathcal{L}_{\mathrm{A}} = \|z_a^{(1)} - z_a^{(2)}\|. \tag{6}$$

To train both the encoders and decoders, we also utilize the maximum log likelihood criterion for the autoencoder loss

$$\mathcal{L}_{\mathrm{AE}} = -[\log p(X^{(1)}|D_{1a}(z_a^{(1)})) + \log p(X^{(2)}|D_{2a}(z_a^{(2)}))].$$

Finally, we combine all losses to a unified learning framework

$$\mathcal{L} = \mathcal{L}_{\mathrm{C}} + \mathcal{L}_{\mathrm{AE}} + \alpha\mathcal{L}_{\mathrm{A}}, \tag{7}$$

where α is the parameter to balance view-alignment and cross-reconstruction.

Intuitively, our final loss is capable of detecting both class anomaly and attribute anomaly. If class anomaly exists, then $z_c^{(1)}$ can not contain the correct information to generate another view $X^{(2)}$, thus leading to high cross-reconstruction loss $\mathcal{L}_{\mathrm{C}}$. If attribute anomaly exists, then $z_a^{(1)}$ and $z_a^{(2)}$ will have weak connection, thus leading to high view-alignment loss $\mathcal{L}_{\mathrm{A}}$.

C. Gumbel-refactored Reconstruction

In unsupervised anomaly detection, since there is no label information to distinguish between normal and anomalous instances, the entire dataset including anomalies are input to the autoencoders. However, anomalies in the dataset could be detrimental to the autoencoder model. On one hand, existing autoencoder is based on accurate reconstruction (e.g., mean square error). Therefore, it can be easily affected by anomalous instances because it tries to fit the anomalous instances as well as normal ones during training. This makes the learned model less discriminative for anomalies. One the other hand,

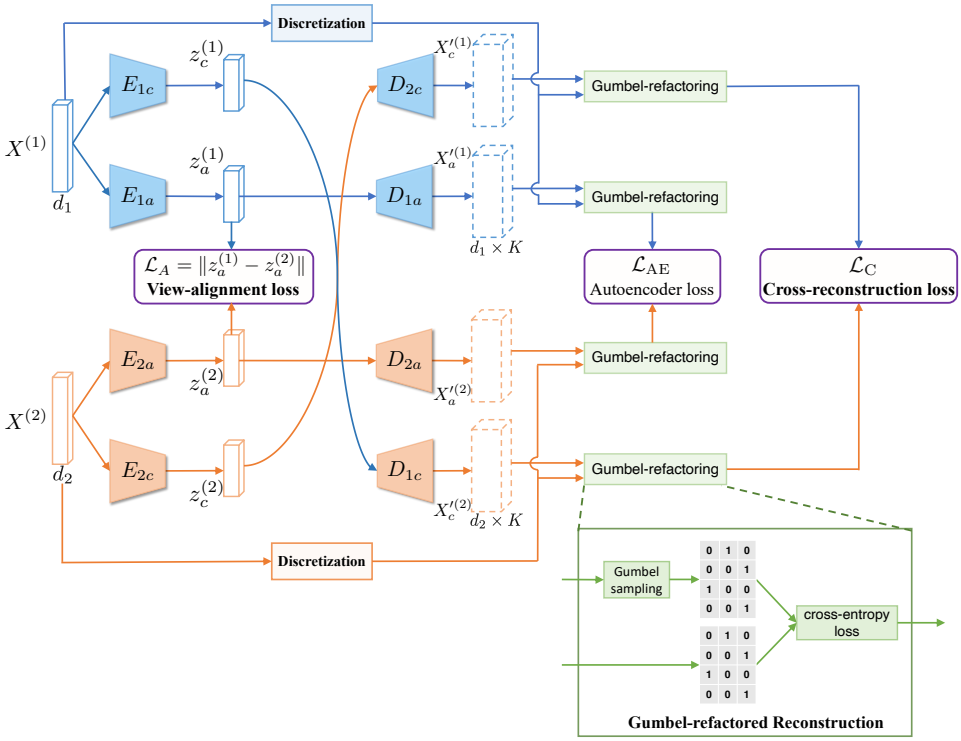


Fig. 2. The proposed CGAEs model. Data from two views are input to different autoencoders to generate the cross-reconstruction loss $\mathcal{L}_{\text{C}}$ (detecting class anomaly), the view-alignment loss $\mathcal{L}_{\text{A}}$ (detecting attribute anomaly), and regular autoencoder loss $\mathcal{L}_{\text{AE}}$. We further propose a Gumbel-refactored reconstruction loss to replace original autoencoder loss to improve the robustness.

the normal instances often contain unavoidable noises due to the data corruption or contamination in the collection process. With accurate reconstruction, the original autoencoder design has the risk of overfitting and modeling the unnecessary noisy details. This further increases the difficulty of distinguishing between normal instances and anomalous ones.

Based on these drawbacks of the original autoencoder, we propose the Gumbel-refactored reconstruction to obtain the robust autoencoders. Ideally, we would like the model to focus on the primary information useful for detecting anomalies and ignore the unnecessary noisy details. In practice, we utilize the idea of discretization and Gumbel sampling. Instead of reconstructing the continuous values of the input X directly, we aim at reconstructing a discrete version of X, denoted as X_d. Specifically, each dimension of the input space is divided into K-bins and X is assigned to one of them to obtain the discretized representation X_d. X_d is transformed into a one-hot vector with 1-of-K bin being 1 and all others being 0. In order to reconstruct the discrete X_d, the decoder is devised to estimate a K-dimensional categorical distribution

$$p(x|z) = \text{Cat}(x|\pi_1, \pi_2, \ldots, \pi_K)\,, \tag{8}$$

where $\text{Cat}(x = i|\pi_1, \pi_2, \ldots, \pi_K) = \pi_i$ and $\sum_{i=1}^{K} = 1$. In order to approximate the one-hot reconstruction X_d' and enable gradient-based optimization for categorical distribution, the Gumbel-Softmax re-parametrization trick [30] is applied. The element i of vector x can be approximated as

$$X_d'[i] = \frac{\exp((\log(\pi_i) + g_i)/\tau)}{\sum_{j=1}^{K} \exp((\log(\pi_j) + g_j)/\tau)} \quad \text{for } i = 1, 2, \ldots, K\,.$$

Here, $g_i \sim \text{Gumbel}(0, 1)$ is the *i.i.d.* samples from the standard Gumbel distribution, τ is the temperature parameter to control the smoothness/discreteness of the distribution, and π is the output of the decoder.

We calculate the Gumbel-refactored reconstruction loss to be the cross-entopy loss between discreterized data X_d and Gumbel reconstruction X_d':

$$\mathcal{L}_{\text{Gum}} = -\sum_{i=1}^{K} X_d[i] \log \left[\frac{\exp((\log(\pi_i) + g_i)/\tau)}{\sum_{j=1}^{K} \exp((\log(\pi_j) + g_j)/\tau)} \right]. \tag{9}$$

In practical implementation, we model all the log likelihood criterion with the Gumbel reconstruction loss, including all log likelihoods in Eq. 3 and Eq. 7. With the discreterized representation and Gumbel-refactored reconstruction, the autoencoders are able to disregard the irrelevant noisy details and encode the essential information for multi-view anomaly detection.

D. Dealing with Multiple Views

Most existing multi-view AD solutions deal with multiple views ($M > 2$) in a pairwise manner, which needs quadratic complexity w.r.t. M. However, our framework can be easily extended to handle multiple views with linear complexity w.r.t M. For the cross-reconstruction module, in order to

generate the reconstruction for view i, we concatenate the $M-1$ latent codes z of all the remaining views. Thus we just need M times computation. For the view-alignment module, we calculate the view-alignment loss Eq. 6 for all possible two-view combinations. Since Eq. 6 is just an Euclidean distance in a low-dimensional space, the computation time is negligible compared with the encoders and decoders. Suppose each encoder and decoder have a fixed time complexity of $O(1)$. Given N instances each including M views, there are $2M$ encoders and $2M$ decoders. Therefore, the overall time complexity is $O(NM)$, which is linear to view number M.

E. Anomaly Scores

To detect two types of multi-view anomalies, we need to compute an anomaly score for each instance. This is done by analyzing the behaviour of normal and abnormal data during the testing phase.

For *normal instances*, since the correlation between views accords with normal training data, it's easier for the model to reconstruct one view from another view. The distance between views of a normal instance can also be easily minimized as they are inherently closer. Both the cross-reconstruction error $\mathcal{L}_{\mathrm{C}}$ and the view-alignment loss $\mathcal{L}_{\mathrm{A}}$ should be small.

For *class anomaly*, since different views belong to different classes, there is inconsistency across multiple views, making it difficult to reconstruct one view from another. This gives rise to a higher cross-reconstruction loss $\mathcal{L}_{\mathrm{C}}$.

For *attribute anomaly*, since the anomalous instances are randomly distributed across different views compared with normal ones, there is a weak connection between the latent codes of different views. This leads to a large view-alignment loss $\mathcal{L}_{\mathrm{A}}$.

Overall, it is supposed that the sum of cross-reconstruction loss $\mathcal{L}_C$ and view-alignment loss $\mathcal{L}_A$ for a normal instance is much smaller than that of an anomalous instance of either *class anomaly* or *attribute anomaly*. Therefore, we compute the anomaly score as follow

$$S = \mathcal{L}_{\mathrm{C}} + \beta \mathcal{L}_{\mathrm{A}}\,, \tag{10}$$

where β is used to balance the two terms.

IV. Experiments

A. Datasets and Settings

1) ***Benchmark Datasets****:* For benchmark datasets, previous algorithms select various datasets from the UCI machine learning repository to generate multi-view anomalies. Different datasets and settings are adopted in the literature. To make a wider range of comparison with previous algorithms, we follow two common settings, i.e. [5] and [9].

In ***Setting 1***, we follow [5] to carry out experiments on three datasets, i.e. Ionosphere, Vowel, and Zoo. Two types of anomalies are generated in the same way as [5]. Specifically, for *Class Anomaly*, two views are obtained by splitting the features into two halves randomly. Then, a certain percentage of multi-view anomalies are generated by swapping the views of two randomly selected instances, which are from different classes. For *Attribute Anomaly*, we randomly choose a sample and replace its feature in all views with random values.

In ***Setting 2***, we follow [9] to conduct experiments on five datasets, i.e. Zoo, Wine, Wdbc, Pima, and Letter. Class anomalies and attribute anomalies are generated according to [9]. In addition, [9] introduces another type of anomaly, namely class-attribute anomaly, which acts like class anomaly in one view and attribute anomaly in another. Our model is capable of detecting this type of anomaly straightforwardly without any modification.

2) ***Real World Datasets****:* For real world datasets, we include NewsM and NewsNG, which are part of the 20 Newsgroup datasets [29]. Each dataset contains three views. To generate anomalous instances, the same anomaly generation procedure as Setting 1 is performed on two randomly chosen views.

3) ***Anomaly Ratios****:* For Setting 1 and real world datasets, following [5], three configurations are considered: (1) 2% as Class Anomaly and 8% as Attribute Anomaly, (2) 5% as Class Anomaly and 5% as Attribute Anomaly, and (3) 8% as Class Anomaly and 2% as Attribute Anomaly. For Setting 2, we follow [9] to generate three configurations for the three anomaly types: (1) 5% for each type, (2) 2%, 5%, 8% for attribute, class and class-attribute anomalies, (3) 8%, 5%, 2% for attribute, class and class-attribute anomalies.

4) ***Implementation and Evaluation Details****:* In the experiments, we empirically set $\alpha = 20, \beta = 800$ to balance different model components and the magnitude of different losses. K is set to 10. For evaluation measure, we adopt the widely-used AUC metric, which is the area under the ROC curve. To reduce the effect of random anomalies generation, we repeat the anomaly generation process for 50 times and report the average results as well as the standard deviations.

B. Comparison with State-of-the-art Methods

We compare with eight state-of-the-art methods: HOrizontal Anomaly Detection (HOAD) [3], Affinity Propagation (AP) [20], MLRA [8], DMOD [6], CRMOD [7], MUVAD [5], LDSR [31], and MODDIS [9]. Among them, HOAD and CC are clustering based methods, DMOD, CRMOD and MLRA rely on transformation on shared latent vector, MUVAD is a nearest neighbor-based method, LDSR and MODDIS are subspace or representation learning related methods.

1) ***Results on Setting 1 and real-word datasets****:* We first show the comparison results on three datasets of Setting 1 and two real world datasets (described in Section IV-A) in Table I. Table I clearly shows that the proposed CGAEs model outperforms all the comparison methods on all datasets and settings with a large margin. This demonstrates the effectiveness of our CGAEs model for unsupervised multi-view anomaly detection.

To be specific, the AUCs of HOAD and CC are inferior because these methods assume that data are well-clustered in each view, which makes them less effective when complex multi-view data distributions violate this assumption. The performance of CRMOD, DMOD and MLRA is also inferior to the proposed CGAEs model, as these methods

TABLE I

AUC VALUES (MEAN±STD) ON FIVE DATASETS. THE NAME IN THE FIRST COLUMN MEANS "DATASET-NUMBEROFVIEWS-CLASSANOMALYRATIO-ATTRIBUTEANOMALYRATIO". FOR EXAMPLE, "ZOO-2-2-8" DENOTES ZOO DATASET, 2 VIEWS, 2% CLASS ANOMALY, AND 8% ATTRIBUTE ANOMALY.

	MUVAD [5]	HOAD [3]	CC [2]	MLRA [8]	DMOD [6]	CRMOD [7]	CGAEs(Our)
Ionosphere-2-2-8	0.834±0.013	0.655±0.051	0.584±0.080	0.831±0.022	0.741±0.023	0.833±0.033	**0.843**±0.025
Ionosphere-2-5-5	0.834±0.018	0.583±0.057	0.566±0.071	0.780±0.022	0.758±0.030	0.743±0.026	**0.857**±0.011
Ionosphere-2-8-2	0.809±0.021	0.547±0.064	0.539±0.052	0.714±0.023	0.766±0.033	0.765±0.030	**0.883**±0.018
Vowel-2-2-8	0.886±0.005	0.349±0.030	0.499±0.025	0.735±0.028	0.851±0.017	0.833±0.021	**0.942**±0.014
Vowel-2-5-5	0.875±0.012	0.389±0.035	0.500±0.033	0.736±0.021	0.733±0.021	0.717±0.027	**0.916**±0.032
Vowel-2-8-2	0.865±0.012	0.454±0.029	0.499±0.028	0.739±0.035	0.619±0.022	0.584±0.031	**0.869**±0.055
Zoo-2-2-8	0.866±0.031	0.491±0.076	0.488±0.066	0.510±0.048	0.823±0.036	0.521±0.071	**0.887**±0.066
Zoo-2-5-5	0.891±0.037	0.474±0.101	0.487±0.087	0.524±0.088	0.786±0.043	0.496±0.102	**0.942**±0.043
Zoo-2-8-2	0.908±0.030	0.525±0.085	0.523±0.105	0.532±0.078	0.727±0.055	0.496±0.079	**0.913**±0.107
NewsM-3-2-8	0.892±0.016	0.498±0.003	0.474±0.037	0.743±0.034	0.873±0.017	0.883±0.017	**0.915**±0.022
NewsM-3-5-5	0.736±0.036	0.548±0.033	0.504±0.042	0.649±0.031	0.716±0.022	0.727±0.023	**0.764**±0.024
NewsM-3-8-2	0.596±0.039	0.558±0.053	0.509±0.039	0.552±0.029	0.569±0.036	0.523±0.036	**0.654**±0.056
NewsNG-3-2-8	0.898±0.019	0.671±0.077	0.450±0.003	0.673±0.023	0.877±0.018	0.887±0.018	**0.906**±0.021
NewsNG-3-5-5	0.751±0.025	0.640±0.054	0.468±0.075	0.631±0.037	0.736±0.021	0.746±0.020	**0.763**±0.037
NewsNG-3-8-2	0.616±0.033	0.536±0.037	0.491±0.017	0.540±0.027	0.581±0.041	0.590±0.043	**0.645**±0.019

rely on the assumption that there exist a linear transform from a commonly shared latent vector to the view-specific vectors. It fails to capture the nonlinearities and impairs the detecting capability for complicated distributions. The most recent method MUVAD is specifically designed for unsupervised multi-view anomaly detection by considering the neighborhood structures of a normal instance across multiple views. However, this neighborhood structure does not always hold in certain complex situations. Instead, our CGAEs utilizes autoencoders to capture the intrinsic intra-view and inter-view information by the proposed cross-reconstruction and view-alignment modules. Therefore, the proposed CGAEs outperforms MUVAD by a big margin.

*2) **Results on Setting 2**:* We further report the comparison results on five datasets of Setting 2 in Table II. Note that Setting 2 is based on [9], which generate three types of anomalies: attribute anomaly, class-attribute anomaly, and class anomaly. Class-attribute anomaly is more like a combination of the other two anomaly types. Thus, without any modification, our CGAEs model can easily detect all three types of anomaly. This is verified by the results in Table II, where CGAEs achieves the best performance in almost all settings. When CGAEs does not achieve the best performance, the margin w.r.t. the best performer is quite small. The second best model is MODDIS [9] which utilizes the neural networks to integrate multi-view data into a latent intact space. However, in MODDIS, there is no reconstruction or cross-view generation like our method, thus being less effective to learn intrinsic codes among views for anomaly detection.

C. Ablation Study

In order to investigate the effectiveness of the designed components (i.e. cross-reconstruction, view-alignment, and Gumbel-refactored reconstruction) in our CGAEs model, we perform various ablation studies.

*1) **Cross-reconstruction ($\mathcal{L}_C$) and View-alignment ($\mathcal{L}_A$)**:* We first investigate the effect of cross-reconstruction ($\mathcal{L}_{\mathrm{C}}$) and view-alignment ($\mathcal{L}_A$). This is done by applying $\mathcal{L}_{\mathrm{C}}$ only, $\mathcal{L}_A$ only, and $\mathcal{L}_A + \mathcal{L}_C$. We experiment on *Vowel* dataset with 5% class anomaly and 5% attribute anomaly. The sorted anomaly scores are visualized in Figure 3. It can be observed that with $\mathcal{L}_A$ only, attribute anomalies show high anomaly scores (i.e. blue bars) while class anomalies are mixed with normal data. With $\mathcal{L}_C$ only, class anomalies show high anomaly scores (i.e. red bars), while attribute anomalies are mixed with normal data. Our final model, $\mathcal{L}_A + \mathcal{L}_C$, can successfully detect both types of anomalies with high scores and achieves the best AUC score. We conclude that each of the proposed component is effective in detecting the specific type of anomaly. Combining the two devised components into a unified model can significantly improve the performance, which further verifies the complementary design.

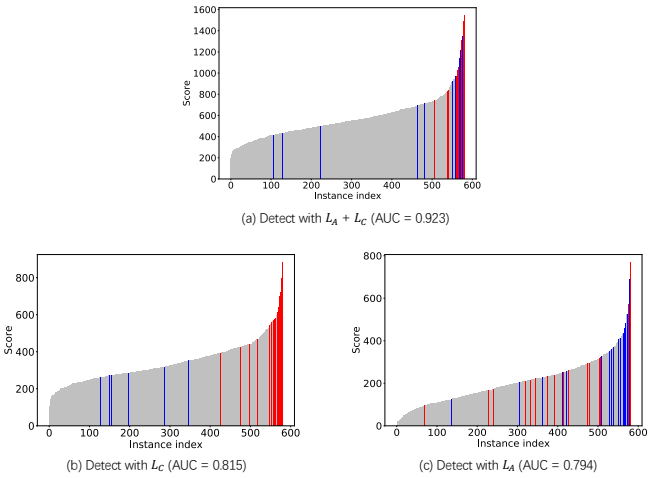


Fig. 3. Sorted anomaly scores for *Vowel* dataset. The instances are sorted from left to right with the vertical bar indicating their anomaly scores. Here, grey bars denote normal data, red bars denote class anomaly, and blue bars denote attribute anomaly. The baseline with neither $\mathcal{L}_{\mathrm{A}}$ nor $\mathcal{L}_{\mathrm{C}}$ obtains AUC$= 0.516$.

*2) **Gumbel-refactored Reconstruction**:* We then investigate the effectiveness of Gumbel-refactored reconstruction. The simple baseline is an autoencoder with mean square error (MSE), which is usually the default choice for autoencoders. Then, a Bernoulli-like distribution is utilized to model the conditional distribution as $p(X_i|X'_i) = {X'_i}^{X_i}(1 - X'_i)^{1-X_i}$, where X_i and X'_i are the i-th dimension of the original input

TABLE II

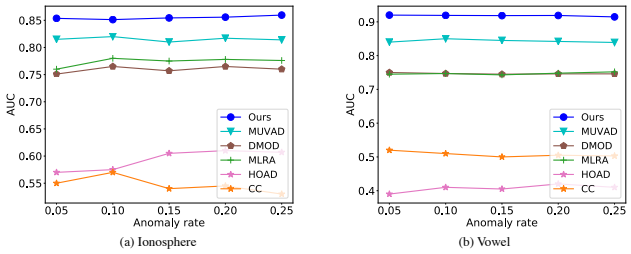
AUC VALUES (MEAN±STD) ON FIVE DATASETS OF SETTING 2. "L2" AND "GAUSS" DENOTE THE L2-NORM AND GAUSSIAN KERNEL UTILIZED AS SIMILARITY METRICS IN HOAD [3] AND AP [20]. WE FOLLOW [9] TO GENERATE THREE TYPES OF ANOMALIES: ATTRIBUTE ANOMALY, CLASS-ATTRIBUTE ANOMALY, AND CLASS ANOMALY, WHOSE RATIOS ARE SHOWN IN THE RATIOS ROW OF THE TABLE.

Dataset	zoo		wine		wdbc		pima		letter	
#Views	v=2	v=3	v=2	v=3	v=2	v=3	v=2	v=3	v=2	v=3
Ratios	5% / 5% / 5%									
$HOAD_{L2}$ [3]	0.55±0.08	0.48±0.08	0.61±0.05	0.61±0.06	0.55±0.06	0.59±0.09	0.59±0.07	0.64±0.05	-	-
$HOAD_{Gauss}$ [3]	0.54±0.08	0.73±0.06	0.59±0.06	0.61±0.06	0.63±0.05	0.54±0.13	0.58±0.04	0.61±0.03	-	-
AP_{L2} [20]	0.77±0.06	0.87±0.04	0.63±0.09	0.65±0.08	0.72±0.04	0.63±0.04	0.59±0.03	0.50±0.04	-	-
AP_{Gauss} [20]	0.79±0.05	0.87±0.03	0.63±0.09	0.66±0.08	0.72±0.04	0.62±0.04	0.59±0.03	0.50±0.04	-	-
MLRA [8]	0.70±0.06	0.71±0.07	0.71±0.05	0.77±0.04	0.71±0.03	0.69±0.05	0.62±0.04	0.64±0.02	0.68±0.01	0.61±0.01
LDSR [31]	0.85±0.03	0.83±0.04	0.78±0.03	0.78±0.03	0.95±0.02	0.92±0.02	0.71±0.02	0.69±0.02	-	-
MODDIS [9]	0.87±0.04	0.87±0.05	0.89±0.03	0.88±0.04	0.94±0.02	0.94±0.01	**0.88**±0.02	**0.87**±0.01	**0.99**±0.01	0.94±0.01
CGAEs(Our)	**0.92**±0.04	**0.90**±0.05	**0.94**±0.02	**0.92**±0.03	**0.98**±0.01	**0.97**±0.02	0.85±0.03	0.85±0.03	0.92±0.02	**0.95**±0.01
Ratios	2% / 5% / 8%									
$HOAD_{L2}$ [3]	0.49±0.09	0.48±0.08	0.57 ±0.06	0.57±0.06	0.56 ±0.09	0.52±0.05	0.48±0.07	0.62±0.05	-	-
$HOAD_{Gauss}$ [3]	0.56±0.07	0.66±0.06	0.53±0.07	0.55 0.06	0.62±0.04	0.53±0.10	0.56±0.04	0.59±0.03	-	-
AP_{L2} [20]	0.77±0.06	0.86±0.04	0.69±0.06	0.69±0.05	0.81±0.03	0.71±0.04	0.63±0.03	0.55±0.03	-	-
AP_{Gauss} [20]	0.80±0.05	**0.87**±0.04	0.69±0.06	0.71±0.05	0.81±0.03	0.70±0.04	0.63±0.03	0.55±0.03	-	-
MLRA [8]	0.67±0.06	0.67±0.06	0.64±0.05	0.70±0.05	0.66±0.03	0.72±0.05	0.59±0.06	0.61±0.02	0.73±0.01	0.63 0.01
LDSR [31]	0.85±0.04	0.84±0.03	0.73±0.05	0.71±0.04	0.91±0.03	0.91±0.03	0.67±0.02	0.64±0.03	-	-
MODDIS [9]	0.86±0.05	0.84±0.05	0.84±0.04	0.80±0.04	0.90±0.02	0.89±0.02	0.82±0.02	**0.80**±0.02	**0.98**±0.01	**0.91**±0.01
CGAEs(Our)	**0.91**±0.06	0.79±0.07	**0.91**±0.04	**0.88**±0.04	**0.96**±0.01	**0.95**±0.01	**0.92**±0.01	**0.80**±0.04	0.96±0.01	**0.91**±0.01
Ratios	8% / 5% / 2%									
$HOAD_{L2}$ [3]	0.50±0.10	0.47±0.11	0.63±0.06	0.64±0.06	0.55±0.07	0.54±0.05	0.53±0.10	0.66±0.09	-	-
$HOAD_{Gauss}$ [3]	0.49±0.11	0.79±0.06	0.62±0.05	0.65±0.05	0.63±0.04	0.57±0.16	0.60±0.04	0.63±0.03	-	-
AP_{L2} [20]	0.77±0.04	0.86±0.04	0.60±0.07	0.65±0.09	0.66±0.04	0.57±0.04	0.54±0.03	0.48±0.04	-	-
AP_{Gauss} [20]	0.79±0.04	0.86±0.04	0.60±0.07	0.66±0.09	0.66±0.04	0.56±0.04	0.54±0.03	0.48±0.04	-	-
MLRA [8]	0.76±0.05	0.75±0.05	0.76±0.05	0.81±0.05	0.63±0.07	0.72±0.06	0.59±0.04	0.64±0.04	0.80±0.01	0.82±0.01
LDSR [31]	0.86±0.04	0.83±0.04	0.81±0.03	0.81±0.04	0.96±0.02	0.95±0.01	0.75±0.02	0.73±0.02	-	-
MODDIS [9]	0.88±0.04	0.88±0.04	0.93±0.03	0.93±0.02	0.98±0.01	0.97±0.01	**0.94**±0.01	**0.94**±0.01	**0.99**±0.01	**0.98**±0.01
CGAEs(Our)	**0.96**±0.03	**0.92**±0.05	**0.96**±0.02	**0.96**±0.01	**0.99**±0.01	**1.00**±0.01	0.92±0.02	0.92±0.01	0.96±0.01	**0.98**±0.00

TABLE III

THE EFFECT OF GUMBEL-REFACTORED RECONSTRUCTION.

	VAE	AE_{MSE}	$AE_{Bernoulli}$	AE_{Gumbel}(Our)
Ion-2-2-8	0.811±0.028	**0.889**±0.040	0.841±0.017	0.843±0.025
Ion-2-5-5	0.802±0.037	0.728±0.020	0.845±0.024	**0.857**±0.011
Ion-2-8-2	0.825±0.028	0.666±0.037	0.817±0.048	**0.883**±0.018
Vowel-2-2-8	0.933±0.018	0.838±0.028	0.846±0.027	**0.942**±0.014
Vowel-2-5-5	0.882±0.046	0.784±0.045	0.820±0.027	**0.916**±0.032
Vowel-2-8-2	0.838±0.045	0.820±0.039	0.792±0.047	**0.868**±0.055
Zoo-2-2-8	0.887±0.054	0.436±0.075	0.691±0.136	**0.886**±0.066
Zoo-2-5-5	0.863±0.084	0.448±0.073	0.706±0.109	**0.942**±0.043
Zoo-2-8-2	0.851±0.088	0.656±0.112	0.833±0.032	**0.913**±0.107



(a) Ionosphere (b) Vowel

Fig. 4. Changes of AUC as anomaly rate increases on *Ionosphere* and *Vowel*.

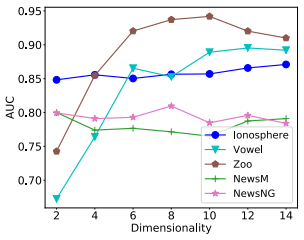


Fig. 5. AUCs with different dimensionalities of the latent code.

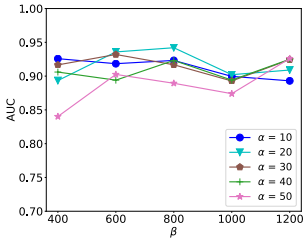


Fig. 6. AUCs with various combinations of α and β on *Zoo* dataset.

X and its reconstruction X'. At last, we utilize varitional autoencoder (VAE) as another baseline. From Table III, we can see that the Gumbel-refactored reconstruction achieves the best performance in 8 out of 9 settings. The simple MSE baseline performs the worst, indicating that accurate and continuous reconstruction is not robust for multi-view anomaly detection. For VAE, it highly relies on the prior distribution (e.g., Gaussian), while the complex multi-view data distribution cannot be easily modeled with a simple prior distribution such as Gaussian. Moreover, in Table III, we observe that the performance of Gumbel-refactored reconstruction tends to have a smaller standard deviation, further demonstrating the robustness of the proposed Gumbel-refactoring module.

*3) **Varying Anomaly Rate**:* To investigate how the anomaly rate affects the performance of different models, we experiment on data polluted by increasing percentages of outliers. Figure 4 shows the changes of AUC on *Ionosphere* and *Vowel* datasets with the anomaly rate increasing from 5% to 25% by a stepsize of 5% for two types of anomaly. Our method outperforms all other methods with a large margin for all anomaly rate values, showing its strong detection power and robustness.

*4) **Dimension of the Latent Space**:* We further investigate the effect of the dimension of the latent space. The latent code in our model has several important functions: 1) view-alignment by comparing the latent codes of different views, 2) cross-reconstruction of one view with the discriminative latent code from another view. Due to the importance of the latent

code, it should be discriminative enough to preserve sufficient information. Therefore, a too small dimension in Figure 5 leads to degraded performance for all datasets. Nevertheless, a too large dimension would also lead to decreased performance on some datasets such as Zoo. We conjecture this is because of the violation of bottleneck design of autoencoder. Based on the analysis, we set the latent code dimension to be 10 in our experiments.

*5) **Balancing Parameters** α **and** β:* Since the computation of $\mathcal{L}_{\mathrm{C}}$ involves larger number of elements than $\mathcal{L}_{\mathrm{A}}$, its magnitude is also larger than $\mathcal{L}_{\mathrm{A}}$. Hence, we set the relatively large α and β to adjust the magnitudes. In specific, we vary α from 10 to 50 and β from 400 to 1200 and show the results in Figure 6. It can be observed that the AUC is quite stable with various combinations of α and β. The best performance is achieved when $\alpha = 20$ and $\beta = 800$. Note that α is set to smaller values compared with β. We explain this from two perspectives: (1) a too large α during training will lead to collapsed latent codes, i.e., $z_a^{(1)} = z_a^{(2)}$; (2) in Eq. 7, both $\mathcal{L}_{\mathrm{AE}}$ and $\mathcal{L}_{\mathrm{A}}$ have the potential effect of dealing with attribute anomaly while in Eq. 10 we only utilize $\mathcal{L}_{\mathrm{A}}$ for attribute anomaly detection, which requires a larger β in Eq. 10.

V. Conclusion

We propose a Cross-aligned and Gumbel-refactored AutoEncoders (CGAEs) model to effectively detect two types of multi-view anomalies. In CGAEs, the cross-reconstruction module is well-devised to detect the class anomaly and the view-alignment module is specifically designed to detect the attribute anomaly. Furthermore, to endue robustness to the proposed CGAEs model, we propose a Gumbel-refactored reconstruction loss to replace the original MSE in autoencoders. Ablation studies verify the effectiveness of each of the designed components, i.e. cross-reconstruction, view-alignment, and Gumbel-refactored reconstruction. Our CGAEs model obtains superior performance on both the benchmark datasets and the real world datasets.

References

[1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[2] A. Y. Liu and D. N. Lam, "Using consensus clustering for multi-view anomaly detection," in *2012 IEEE Symposium on Security and Privacy Workshops*. IEEE, 2012, pp. 117–124.

[3] J. Gao, W. Fan, D. Turaga, S. Parthasarathy, and J. Han, "A spectral framework for detecting inconsistency across multi-source object relationships," in *International Conference on Data Mining*, 2011.

[4] K. Duh, C. man Au Yeung, T. Iwata, and M. Nagata, "Managing information disparity in multilingual document collections," *ACM Trans. Speech Lang. Process.*, vol. 10, pp. 1:1–1:28, 2013.

[5] X.-R. Sheng, D.-C. Zhan, S. Lu, and Y. Jiang, "Multi-view anomaly detection: Neighborhood in locality matters," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019.

[6] H. Zhao and Y. Fu, "Dual-regularized multi-view outlier detection," in *International Joint Conference on Artificial Intelligence*, 2015.

[7] H. Zhao, H. Liu, Z. Ding, and Y. Fu, "Consensus regularized multi-view outlier detection," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 236–248, 2017.

[8] S. Li, M. Shao, and Y. Fu, "Multi-view low-rank analysis for outlier detection," in *SIAM International Conference on Data Mining*, 2015.

[9] Y.-X. Ji, L. Huang, H.-P. He, C.-D. Wang, G. Xie, W. Shi, and K.-Y. Lin, "Multi-view outlier detection in deep intact space," in *International Conference on Data Mining*, 2019.

[10] T. Iwata and M. Yamada, "Multi-view anomaly detection via robust probabilistic latent variable models," in *Advances In Neural Information Processing Systems*, 2016, pp. 1136–1144.

[11] H. Wang, T. Chen, H. Wang, X. Shao, and P. Su, "Fuzzy clustering based anomaly detection for distributed multi-view data," in *International Conference on Fuzzy Systems*, 2018.

[12] H. Wang, P. Su, M. Zhao, H. Wang, and G. Li, "Multi-view group anomaly detection," in *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, 2018, pp. 277–286.

[13] K. Singh and S. Upadhyaya, "Outlier detection: applications and techniques," *International Journal of Computer Science Issues (IJCSI)*, vol. 9, no. 1, p. 307, 2012.

[14] Z. Yang, T. Zhang, I. S. Bozchalooi, and E. Darve, "Memory augmented generative adversarial networks for anomaly detection," *arXiv preprint arXiv:2002.02669*, 2020.

[15] Z. Wang and C. Lan, "Towards a hierarchical bayesian model of multi-view anomaly detection," in *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, 2020.

[16] E. Šabić, D. Keeley, B. Henderson, and S. Nannemann, "Healthcare and anomaly detection: using machine learning to predict anomalies in heart rate data," *AI & SOCIETY*, pp. 1–10, 2020.

[17] T. Pourhabibi, K.-L. Ong, B. H. Kam, and Y. L. Boo, "Fraud detection: A systematic literature review of graph-based anomaly detection approaches," *Decision Support Systems*, vol. 133, 2020.

[18] S. Das, B. L. Matthews, A. N. Srivastava, and N. C. Oza, "Multiple kernel learning for heterogeneous anomaly detection: algorithm and aviation safety case study," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 47–56.

[19] V. P. Janeja and R. Palanisamy, "Multi-domain anomaly detection in spatial datasets," *Knowledge and information systems*, vol. 36, no. 3, pp. 749–788, 2013.

[20] A. Marcos Alvarez, M. Yamada, A. Kimura, and T. Iwata, "Clustering-based anomaly detection in multi-view data," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*. ACM, 2013, pp. 1545–1548.

[21] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in neural information processing systems*, 2002, pp. 849–856.

[22] B. J. Frey and D. Dueck, "Mixture modeling by affinity propagation," in *Advances in neural information processing systems*, 2006, pp. 379–386.

[23] D. Hou, Y. Cong, G. Sun, J. Dong, J. Li, and K. Li, "Fast multi-view outlier detection via deep encoder," *IEEE Transactions on Big Data*, 2020.

[24] B. Zong, Q. Song, M. R. Min, W. Cheng, C. Lumezanu, D. Cho, and H. Chen, "Deep autoencoding gaussian mixture model for unsupervised anomaly detection," in *ICLR*, 2018.

[25] D. Gong, L. Liu, V. Le, B. Saha, M. R. Mansour, S. Venkatesh, and A. v. d. Hengel, "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *CVPR*, 2019.

[26] A. Borghesi, A. Bartolini, M. Lombardi, M. Milano, and L. Benini, "Anomaly detection using autoencoders in high performance computing systems," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 9428–9433.

[27] C. Wang, B. Wang, H. Liu, and H. Qu, "Anomaly detection for industrial control system based on autoencoder neural network," *Wireless Communications and Mobile Computing*, 2020.

[28] K. Deepak, G. Srivathsan, S. Roshan, and S. Chandrakala, "Deep multi-view representation learning for video anomaly detection using spatiotemporal autoencoders," *Circuits, Systems, and Signal Processing*, pp. 1–17, 2020.

[29] G. Bisson and C. Grimal, "Co-clustering of multi-view datasets: a parallelizable approach," in *2012 IEEE 12th International Conference on Data Mining*. IEEE, 2012, pp. 828–833.

[30] E. Jang, S. Gu, and B. Poole, "Categorical reparameterization with gumbel-softmax," *arXiv preprint arXiv:1611.01144*, 2016.

[31] K. Li, S. Li, Z. Ding, W. Zhang, and Y. Fu, "Latent discriminant subspace representations for multi-view outlier detection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.