# Conditioning Latent-Space Clusters for Real-World Anomaly Classification

Daniel Bogdoll[†‡∗], Svetlana Pavlitska[†‡∗], Simon Klaus[‡∗], and J. Marius Zöllner[†‡]

[†]FZI Research Center for Information Technology, Germany

bogdoll@fzi.de

[‡]Karlsruhe Institute of Technology, Germany

*Abstract*—Anomalies in the domain of autonomous driving are a major hindrance to the large-scale deployment of autonomous vehicles. In this work, we focus on high-resolution camera data from urban scenes that include anomalies of various types and sizes. Based on a Variational Autoencoder, we condition its latent space to classify samples as either normal data or anomalies. In order to emphasize especially small anomalies, we perform experiments where we provide the VAE with a discrepancy map as an additional input, evaluating its impact on the detection performance. Our method separates normal data and anomalies into isolated clusters while still reconstructing high-quality images, leading to meaningful latent representations.

*Index Terms*—anomaly, corner case, vision, autonomous driving, VAE, latent space, cluster

## I. Introduction

Developing autonomous vehicles and deploying them to large Operational Design Domains (ODD) poses a significant challenge, especially with respect to the long tail of unexpected or unfamiliar objects. Although perception systems of autonomous vehicles are able to detect known classes reasonably well nowadays, they still need to be aware of situations where they encounter the unknown. As deep neural networks (DNN) tend to predict false positives with high uncertainty, both false negatives and false positives are of interest. While often lidar sensors are used in production systems, here we focus on a purely camera-based setup, as those introduce their own set of issues. We focus on detecting anomalies in high-resolution road images from mostly urban scenes based on the latent space of a Variational Autoencoder (VAE). Utilizing primarily normal but also abnormal data during training, the data is being fit on two prior distributions. This way, the VAE is conditioned to build two separate clusters in the latent space, one for normal samples and one for anomalies. During test time, distance measures can be used to detect anomalies. We used multiple datasets to define normality and anomalies during training and evaluation. The work is structured as follows: In Section II, we introduce related work from the field of anomaly detection, with a focus on Variational Autoencoders. In Section III, we introduce our approach, including our VAE architecture. In Section IV, we highlight our experimental setup and demonstrate our results. Finally, we conclude this work in Section V. More information can be found in [1].

∗ These authors contributed equally



Fig. 1: Our VAE-based method for real-world anomaly classification, which separates normal and abnormal data in its latent space. Discrepancy images as additional inputs also emphasize small unknown objects, here *a cat*.

## II. Related Work

In autonomous driving, detecting anomalies is of utmost importance to scale existing systems, which operate in small Operational Design Domains, as infrequent events occur more often for a growing vehicle fleet that utilizes the same software system [2]. Based on common corner case systematizations [3]–[5], we are especially interested in the *object* and *scene* levels, which describe unknown objects or, more generally, unexpected patterns in an input sample. In this section, we introduce current approaches for such detections, also known as outliers or out-of-distribution (OOD) samples.

As one of the early approaches, Ruff et al. [6] trained a deep neural network in order to compute a hypersphere representing an approximation of normality. A binary classification can be computed based on the distance of new samples to the hypersphere. However, in the real world, normality is represented in the form of many "heterogeneous semantic labels" [7], which leads to a weak decision boundary. Hendrycks et al. proposed an "outlier exposure" objective [8], utilizing curated anomaly data for training, leading to a more uniform softmax distribution for anomalies, which was later adopted by Papadopoulos et al. [9]. Some approaches utilize the softmax confidence in order to detect anomalies [10], [11]. However, this can lead to false positives since the softmax activation function is sensitive to changes in the input. Thus,

Liu et al. proposed an energy-based approach [12], showing an improved alignment to the density of the input samples.

Based on a GAN, Nitsch et al. generated virtual anomalies for an improved decision boundary [13]. Similarly, Grcic et al. used normalizing flows for the same task [14], outperforming most other methods [15]. Going one step further, Du et al. introduced Virtual Outlier Synthesis (VOS), which generates synthetic anomalies in the latent space [16]. In the field of semantic segmentation, Cen et al. included unknown objects in their class list, leading to an open-world segmentation approach based on Euclidean distances between feature vectors [17]. Di Biase integrated model uncertainty into their system in order to reduce wrong classifications [18].

As we have shown, there exist many different methods to detect anomalies. Since our work is based on VAEs, we now provide a detailed overview of methods based on encodings.

### A. Encoding-based Anomaly Detection

Breitenstein et al. have categorized anomaly detection in the domain of autonomous driving into five different categories: Reconstruction, Prediction, Generative, Confidence Score, Feature Extraction [19]. In this work, we examine the properties of Variational Autoencoders in order to classify image samples as anomalies. VAEs can be utilized for dense anomaly detections, which fall under either the *Reconstruction* or *Generative category*, and anomaly classifications, which can be categorized as *Feature Extraction*. Methods from the first category are based on the assumption that the utilized training data defines normality, leading to failed reconstructions given samples that include parts that were not included in the training data. Methods from the second category, which take a look at the latent space, assume that latent representations from normal and OOD samples have a sufficient difference.

**Reconstructive and Generative.** A well-trained VAE will reconstruct unseen anomalies [20], which is why specialized methods were developed for anomaly detection. Utilizing a Generative Adversarial Network (GAN), in which both the generator and the discriminator are implemented as AE, Vu et al. designed a network where the discriminator learns to reconstruct normal data while failing to do so when presented with OOD data [21]. Utilizing the reconstruction probability, An et al. go beyond utilizing the direct reconstruction error, which is incapable of incorporating high-level structures [22]. Among others, Munjal et al. introduced an adversarial loss in order to address this issue [23]. However, this method is not effective for high-complexity scenes. On a similar note, Somepalli et al. proposed to minimize the Wasserstein distance and include a latent space regularization, which led to better reconstructions of normal data and worse ones for anomalies. On the other hand, Bolte et al. [24] proposed a multi-stage approach, where an Autoencoder was used for image prediction. Based on this, an engineered approach followed, which included prediction errors, pixel classification, and distance weighting. Similarly, Amini et al. [25] proposed a pixel-wise uncertainty for reconstructions with a VAE. It is also possible to combine both methods which are based



Fig. 2: We used the Cityscapes [29] and Fishyscapes [30] (normal) datasets as normality (left) and the RoadAnomaly21 [31], Fishyscapes (anomalies), and Lost and Found [32] datasets with anomalies (right). Reprinted from [1].

on reconstructions and feature extraction. Abati et al. [26] utilize reconstruction error and latent features with a low log-likelihood in order to detect anomalies. Similarly, Wang et al. [27] use a discrete latent space, where a model learns the distribution. Reconstructions are then based upon a re-sampled latent space and used for anomaly detection. Park et al. [28] proposed a memory module where prototypes of normality are stored. Later, a reconstruction-based approach, using these memory items, is used for anomaly detection.

**Feature Extraction.** As some previous techniques already utilized feature extraction partly, this section highlights works focusing on this technique. Wurst et al. [33] utilized a triplet-based Autoencoder, enforcing similarity in the latent space, to detect unusual traffic scenes. Similarly, Harmening et al. [34] uses clusters in the latent space generated by an Autoencoder to detect novel scenarios. For more complex data, Sundar et al. [35] developed a method to divide datasets into smaller subsets. Based on these, multiple VAEs are trained to generate the latent space. For detection, they utilize all trained VAEs and detect high sensitivity. Akcay et al. [36] compared latent representations of image reconstructions and the original input to detect anomalies. Chalapathy at al. [37] utilize a one-class classification objective based on features learned by a VAE, where they focus on generating features that are designed for the task of anomaly detection [38]. Park et al. [39] utilize rate-distortion theory in order to compute anomaly scores, only using the encoding part of a VAE. The work of Liu et al. [40] is based on attention maps for every element of the latent vector, where they compute differences to the learned normality, leading to an attention map that highlights anomalies in an image. Finally, Dilokthanakul et al. [41] proposed a VAE which uses a mixture of Gaussians as prior, assuming multiple distributions in the training data, which led to a better separation of classes in the latent space.

While many of the presented approaches work with simple datasets, in our work, we are interested in high-resolution images [42], [43] with anomalies in urban road scenarios [44]. Here, the challenge arises that anomalies often only occupy small regions of an image, which makes classification harder, as normality is represented by highly complex training data. Our approach evaluates whether an auxiliary input that highlights even small anomalies in the image space, combined

(a) Overall architecture of the VAE.  (b) ResBlock

Fig. 3: Overall architecture of the deployed VAE (left) and the components of the ResBlock (right). Adapted from [1]



Fig. 4: Discrepancy images for a Cityscapes image containing an object of the rare but normal class *bus*. The original approach by Lis et al. [49] (middle) leads to higher anomaly scores. The proposed frequency-based approach (right) leads to lower anomaly scores. Reprinted from [1].

with a conditioned latent space, allows for the classification of anomalies in such datasets.

## III. METHOD

In the following, we describe our anomaly detection method, which conditions the latent space of a VAE to enforce separations of clusters corresponding to anomalous and normal data.

### A. VAE Architecture

We use a conditioned latent space variational autoencoder (CL-VAE) [45]. Our architecture roughly follows Hou et al. [46]. However, we utilize residual blocks, as shown in Figure 3a[1], as those are easier to train. Furthermore, we inserted an additional pooling layer and adjusted the residual block by replacing the exponential linear unit (ELU) activation function with the randomized leaky rectified linear units (RRelu) as proposed by Xu et al. [47], see Figure 3b. The VAE was trained with the CL-VAE ELBO loss. We performed experiments with two auxiliary losses to enforce the separation of normal and anomalous data. First, the distance loss $\mathcal{L}_{distance}$ aims at maximizing the distance between the two means to push the clusters away from each other:

$$\mathcal{L}_{distance} = -\|\mu_1, \mu_2\|_1 = -|\mu_1 - \mu_2| \qquad (1)$$

Second, we use $\mathcal{L}_i$ proposed by Yang et al. [48] to maximize the distance between each data point and the cluster mean given by the k-means algorithm: $\mathcal{L}_i = (\mu_i - z_i)^2$. The overall *cluster loss* is then given by:

$$\mathcal{L}_{cluster} = \frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_i \qquad (2)$$

We also incorporate the feature perceptual loss [46] using a pre-trained backbone to enforce reconstruction quality. This concept ensures meaningful latent representations of the input samples, which can be used for downstream tasks. Discrepancy images passed as the fourth channel were not considered as the backbone was pre-trained on RGB images.

### B. Generation of Discrepancy Images

We use discrepancy images, highlighting areas with anomalous objects' locations, as an additional input to a VAE as a fourth channel. Following the method proposed by Lis et al. [49], we first create a semantic segmentation prediction

[1]Implementation inspired by LukeDitria/CNN-VAE

for a given image. A GAN then tries to recreate the original image from this semantic segmentation image. Finally, the discrepancy network is used to generate the discrepancy image by comparing this recreated image to its counterpart. The discrepancy network comprises three streams: a pre-trained CNN extract features from an original and a resynthesized image, and a custom CNN extracts features from a semantic segmentation map. The extracted features pass through a decoder which outputs the resulting discrepancy image.

In the original approach by Lis et al. [49], the discrepancy detector is trained on the dataset of normal data with altered labels. In particular, labels of some randomly selected objects are replaced with random class labels, thus creating synthetic anomalies. However, due to natural class imbalance, the model learned to classify objects of rare classes as anomalies because randomly choosing a replacement class makes them occur more frequently as an anomaly replacement class than a normal class. To mitigate this issue, we propose the *frequence-based label replacement* as shown in Figure 4. To create a synthetic anomaly dataset for training, rare classes, i.e., those which occur less frequently in a dataset of normal data, are chosen as frequently as a replacement as common ones.

## IV. EVALUATION

In the following, we describe the evaluation of our anomaly detection method. First, we provide details on our experimental setup, followed by several analyses.

### A. Experimental Setting

**Training Data:** We utilized three datasets to train the VAE. *Cityscapes* [29] was used to to represent normal data and both *LostAndFound* [32] and *RoadAnomaly21* from the *SegmentMeIfYouCan* benchmark [31] were used to represent anomalous data. Samples from these datasets can be found in Figure 2. For Cityscapes, we used the pre-defined train-val-test split. The LostAndFound dataset was filtered as follows: We deleted images with less than 3,000 anomalous pixels per image and images containing children, as those are considered normal in Cityscapes. We have selected only a few images with different anomalies from each scene to avoid overfitting. The resulting filtered dataset thus contained 172 train, 99 validation, and 64 test images. Finally, all 110 images from the RoadAnomaly21 dataset were split according to the 70:20:10 rule. For training, all images were downsampled to $256 \times 256$.

Fig. 5: Distribution of mean anomaly scores in the discrepancy maps generated for the Cityscapes test set, comparing the original approach by Lis et al. [49] (blue) to our frequency-based label replacement (orange). Reprinted from [1].



Fig. 6: ROC curves for anomaly detection using the discrepancy maps generated with the proposed frequency-based label replacement: LostAndFound (left) and RoadAnomaly (right) test data. Reprinted from [1].

**Test Data:** For evaluation, the test data from LostAndFound and RoadAnomaly21 datasets were used, which were split as described above. We also used *FS Static* images from the *Fishyscapes* dataset [30]. Because of the small dataset size, it was only used at the test stage. Just 30 images are publicly available, 10 with normal and 20 with anomalous data.

**Models and Training:** Following the approach proposed by Lis et al. [49], we used a pre-trained PSPNet [50] with a pre-trained ResNet backbone [51] to predict semantic segmentation masks for input images and a pre-trained pix2pixHD model [52] for image resynthesis. The discrepancy module included a pre-trained VGG [53] for feature extraction. The VAE was trained for 100 epochs using the ADAM optimizer [54] with a learning rate of 1e-4 and a batch size of 12. The learning rate decreased linearly during training. All trainings were performed on an Nvidia GeForce RTX 3090.

### B. Impact of Frequency-based Label Replacement

In our discrepancy module, we used Cityscapes as a normal dataset where no anomalies should appear. We have analyzed the average pixel-wise anomaly score in generated grayscale discrepancy images, where 0 corresponds to normal and 1 to anomalous data. Ideally, all discrepancy scores should be zero, as no anomalies exist in the data. As Figure 5 demonstrates, the average pixel value in discrepancy maps is lower for the proposed frequency-based label replacement variant than the random-class approach proposed by Lis et al. [49]. Furthermore, a visual comparison of the resulting discrepancy images as shown in Figure 4 confirms that our frequency-based class selection results in lower anomaly scores for normal classes. Furthermore, we evaluated the impact of the frequency-based class selection on LostAndFound and RoadAnomaly using the anomaly detector from Lis et al. [49]. Figure 6 shows that our approach leads to improved classifications for RoadAnomaly dataset but worse results for LostAndFound.

### C. VAE Reconstruction Performance

We evaluated the impact of two hyperparameters on the reconstruction performance: The size of the latent space and the $\beta$ parameter of the KL divergence. We used the Fréchet Inception Distance (FID) [55] to measure the quality of the reconstructions. Figure 8 shows that both a larger latent space and smaller $\beta$ lead to more accurate reconstructions. Finally, we evaluated the impact of the feature perceptual loss on the reconstruction quality. Figure 7 shows that using the feature perceptual loss results in less blurry images.



Fig. 7: Image reconstruction by a VAE with the latent space size of $512 \times 4 \times 4$ and a $\beta = 0.01$ trained with (right) and without (left) the feature perceptual loss. Reprinted from [1].

### D. Impact of Discrepency Image

To evaluate the effect of the discrepancy maps, we first calculated the mean pixel scores for both normal and abnormal data. We found that the score is much higher for images including anomalies than those without. However, an ablation study without the input revealed that the discrepancy map had little effect on the structure of the latent space, especially high-dimensional latent states.

### E. Anomaly Classification via Clustering

To classify an image as normal or anomalous during evaluation, K-Means clustering of the latent space of the trained VAE is performed. We used PCA to visualize the distribution of inputs in the latent space. Our experiments have shown that the larger size of the latent space improves the reconstruction strength of the VAE, as shown above, and the clustering in the latent space. A large latent space size $512 \times 4 \times 4$ led to better results than small ones like $64 \times 4 \times 4$ (see Figure 9).

A quantitative analysis of cluster assignments for different $\beta$ values, as shown in Table I, has revealed that smaller $\beta$

Fig. 8: Image reconstructions for different $\beta$ values (top) and latent map sizes (bottom) of a VAE with the latent feature map of size $z \times 4 \times 4$. Average FID and MSE values were measured on the Cityscapes test dataset. Adapted from [1].



(a) Groundtruth: Latent space size of $64 \times 4 \times 4$

(b) Utilized datasets for VAE with $\beta = 0.01$

(c) Groundtruth: Latent space size of $512 \times 4 \times 4$.

(d) Groundtruth: VAE with $\beta = 0.01$

(e) Clustering: Latent space of size $512 \times 4 \times 4$.

(f) Clustering: VAE with $\beta = 0.01$

Fig. 9: Impact of the dimensionality (left) and $\beta = 0.01$ (right) on clustering the latent space of a VAE. Adapted from [1].

values lead to lower false positive rates. On the right side of Figure 9, it can be seen that for $\beta = 0.01$, the proposed approach can detect most anomalous data, i.e., data points corresponding to three anomaly datasets. Adding the previously described cluster loss in Equation 2 did not help to reduce

the number of false positives. Furthermore, the distance loss from Equation 1 significantly increased the distance between the clusters. However, this latent space structure is unsuitable for cluster separation for normal and anomalous data [1].

| $\beta$ | 1 | 0.1 | 0.01 | 0.001 |
|---|---|---|---|---|
| FPR | 0,4332 | 0,3231 | 0,3557 | 0,4065 |
| TPR | 0,9894 | 0,9681 | 1 | 1 |

TABLE I: False and true positive rate for anomaly classification using a VAE with latent space for different $\beta$ values.

## V. CONCLUSION

In this work, we have presented an approach to detect image samples containing anomalies based on the latent space of a Variational Autoencoder. The latent space was conditioned in a way to create individual clusters for those categories, which allowed for the detection of anomalies during inference. We could show, that our model is even able to detect small anomalies from datasets without a domain shift compared to the training data. However, similar to other anomaly detection approaches [16], our method still produces many false positives. We have performed experiments with different components, such as a distance loss, a cluster loss, or an additional discrepancy map as the input, evaluating their impact on the performance of the model. While high false-positive rates are not suitable for production systems, our approach can be utilized for an active learning system, where a human oracle can choose relevant frames from a pre-selection, based on the detection results from our method.

## VI. ACKNOWLEDGMENT

REFERENCES

[1] S. Klaus, "Anomaly Detection in the Latent Space of VAEs," Bachelor Thesis, Karlsruhe Institute of Technology (KIT), 2022.

[2] S. Houben *et al.*, "Inspect, understand, overcome: A survey of practical methods for AI safety," in *Deep Neural Networks and Data for Automated Driving: Robustness, Uncertainty Quantification, and Insights Towards Safety*. Springer, 2022.

[3] J. Breitenstein *et al.*, "Systematization of Corner Cases for Visual Perception in Automated Driving," in *Intelligent Vehicles Symposium (IV)*, 2020.

[4] F. Heidecker *et al.*, "An Application-Driven Conceptualization of Corner Cases for Perception in Highly Automated Driving," in *Intelligent Vehicles Symposium (IV)*, 2021.

[5] D. Bogdoll *et al.*, "Description of Corner Cases in Automated Driving: Goals and Challenges," in *International Conference on Computer Vision (ICCV) - Workshops*, 2021.

[6] L. Ruff *et al.*, "Deep One-Class Classification," in *International Conference on Machine Learning (ICML)*, 2018.

[7] J. Park *et al.*, "What is Wrong with One-Class Anomaly Detection?" in *International Conference on Learning Representations (ICLR) - Workshops*, 2021.

[8] D. Hendrycks *et al.*, "Deep Anomaly Detection with Outlier Exposure," in *International Conference on Learning Representations (ICLR)*, 2019.

[9] A.-A. Papadopoulos *et al.*, "Outlier exposure with confidence control for out-of-distribution detection," *Neurocomputing*, vol. 441, 2021.

[10] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *International Conference on Learning Representations (ICLR)*, 2017.

[11] S. Liang *et al.*, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *International Conference on Learning Representations (ICLR)*, 2018.

[12] W. Liu *et al.*, "Energy-based Out-of-distribution Detection," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2020.

[13] J. Nitsch *et al.*, "Out-of-Distribution Detection for Automotive Perception," in *International Conference on Intelligent Transportation Systems (ITSC)*, 2021.

[14] M. Grcić *et al.*, "Dense anomaly detection by robust learning on synthetic negative data," *arXiv:2112.12833*, 2021.

[15] D. Bogdoll *et al.*, "Anomaly Detection in Autonomous Driving: A Survey," in *Conference on Computer Vision and Pattern Recognition (CVPR) - Workshops*, 2022.

[16] X. Du *et al.*, "VOS: Learning What You Don't Know by Virtual Outlier Synthesis," *International Conference on Learning Representations (ICLR)*, 2022.

[17] J. Cen *et al.*, "Deep Metric Learning for Open World Semantic Segmentation," in *International Conference on Computer Vision (ICCV)*, 2021.

[18] G. Di Biase *et al.*, "Pixel-Wise Anomaly Detection in Complex Driving Scenes," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021.

[19] J. Breitenstein *et al.*, "Corner Cases for Visual Perception in Automated Driving: Some Guidance on Detection Approaches," *arXiv:2102.05897*, 2021.

[20] D. Bogdoll *et al.*, "Compressing Sensor Data for Remote Assistance of Autonomous Vehicles using Deep Generative Models," *Advances in Neural Information Processing Systems (NeurIPS) - Workshops*, 2021.

[21] H. S. Vu *et al.*, "Anomaly Detection with Adversarial Dual Autoencoders," *arXiv:1902.06924*, 2019.

[22] J. An and S. Cho, "Variational Autoencoder based Anomaly Detection using Reconstruction Probability," in *SNU Data Mining Center - Special Lecture on IE*, 2015.

[23] P. Munjal *et al.*, "Implicit Discriminator in Variational Autoencoder," in *International Joint Conference on Neural Networks (IJCNN)*, 2020.

[24] J.-A. Bolte *et al.*, "Towards Corner Case Detection for Autonomous Driving," in *Intelligent Vehicles Symposium (IV)*, 2019.

[25] A. Amini *et al.*, "Variational Autoencoder for End-to-End Control of Autonomous Driving with Novelty Detection and Training De-biasing," in *International Conference on Intelligent Robots and Systems (IROS)*, 2018.

[26] D. Abati *et al.*, "Latent Space Autoregression for Novelty Detection," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[27] L. Wang *et al.*, "Image Anomaly Detection Using Normal Data Only by Latent Space Resampling," *Applied Sciences*, vol. 10, 2020.

[28] H. Park *et al.*, "Learning Memory-Guided Normality for Anomaly Detection," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[29] M. Cordts *et al.*, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[30] H. Blum *et al.*, "The Fishyscapes Benchmark: Measuring Blind Spots in Semantic Segmentation," *International Journal of Computer Vision*, vol. 129, 2021.

[31] R. Chan *et al.*, "SegmentMeIfYouCan: A Benchmark for Anomaly Segmentation," *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

[32] P. Pinggera *et al.*, "Lost and Found: detecting small road hazards for self-driving vehicles," in *International Conference on Intelligent Robots and Systems (IROS)*, 2016.

[33] J. Wurst *et al.*, "Novelty Detection and Analysis of Traffic Scenario Infrastructures in the Latent Space of a Vision Transformer-Based Triplet Autoencoder," in *Intelligent Vehicles Symposium (IV)*, 2021.

[34] N. Harmening *et al.*, "Deep Representation Learning and Clustering of Traffic Scenarios," in *International Conference on Machine Learning (ICML) - Workshops*, 2020.

[35] V. K. Sundar *et al.*, "Out-of-Distribution Detection in Multi-Label Datasets using Latent Space of $\beta$-VAE," in *Symposium on Security and Privacy (S&P) - Workshops*, 2020.

[36] S. Akcay *et al.*, "GANomaly: Semi-supervised Anomaly Detection via Adversarial Training," in *Asian Conference on Computer Vision (ACCV)*, Cham, 2018.

[37] R. Chalapathy *et al.*, "Anomaly Detection using One-Class Neural Networks," *arXiv:1802.06360*, 2019.

[38] S. M. Erfani *et al.*, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," *Pattern Recognition*, vol. 58, 2016.

[39] S. Park *et al.*, "Interpreting Rate-Distortion of Variational Autoencoder and Using Model Uncertainty for Anomaly Detection," *Annals of Mathematics and Artificial Intelligence*, vol. 90, 2020.

[40] W. Liu *et al.*, "Towards Visually Explaining Variational Autoencoders," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

[41] N. Dilokthanakul *et al.*, "Deep Unsupervised Clustering with Gaussian Mixture Variational Autoencoders," *arXiv:1611.02648*, 2017.

[42] D. Bogdoll *et al.*, "AD-Datasets: A Meta-Collection of Data Sets for Autonomous Driving," in *International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS)*, 2022.

[43] ——, "Impact, Attention, Influence: Early Assessment of Autonomous Driving Datasets," in *International Conference on Control and Robotics Engineering (ICCRE)*, 2023.

[44] ——, "Perception Datasets for Anomaly Detection in Autonomous Driving: A Survey," in *Intelligent Vehicles Symposium (IV)*, 2023.

[45] E. Norlander and A. Sopasakis, "Latent space conditioning for improved classification and anomaly detection," *arXiv:1911.10599*, 2019.

[46] X. Hou *et al.*, "Deep Feature Consistent Variational Autoencoder," in *Winter Conference on Applications of Computer Vision (WACV)*, 2017.

[47] B. Xu *et al.*, "Empirical Evaluation of Rectified Activations in Convolutional Network," *arXiv:1505.00853*, 2015.

[48] B. Yang *et al.*, "Towards K-means-friendly Spaces: Simultaneous Deep Learning and Clustering," in *International Conference on Machine Learning (ICML)*, 2017.

[49] K. Lis *et al.*, "Detecting the Unexpected via Image Resynthesis," in *International Conference on Computer Vision (ICCV)*, 2019.

[50] H. Zhao *et al.*, "Pyramid Scene Parsing Network," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[51] K. He *et al.*, "Deep Residual Learning for Image Recognition," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[52] T. Wang *et al.*, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[53] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations (ICLR)*, 2015.

[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[55] M. Heusel *et al.*, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.