This WACV 2022 paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

A Semi-supervised Generalized VAE Framework for Abnormality Detection using One-Class Classification

Renuka Sharma[1,2], Satvik Mashkaria[1], and Suyash P. Awate[1]
[1]Computer Science and Engineering Department, Indian Institute of Technology Bombay, Mumbai.
[2]IITB-Monash Research Academy, Mumbai.

Abstract

Abnormality detection is a one-class classification (OCC) problem where the methods learn either a generative model of the inlier class (e.g., in the variants of kernel principal component analysis) or a decision boundary to encapsulate the inlier class (e.g., in the one-class variants of the support vector machine). Learning schemes for OCC typically train on data solely from the inlier class, but some recent OCC methods have proposed semi-supervised extensions that also leverage a small amount of training data from outlier classes. Other recent methods extend existing principles to employ deep neural network (DNN) models for learning (for the inlier class) either latent-space distributions or autoencoders, but not *both. We propose a* semi-supervised variational *formulation, leveraging* generalized-Gaussian *(GG) models leading to data-adaptive,* robust, *and* uncertainty-aware *distribution modeling in both latent space and image space. We propose a* reparameterization *for sampling from the latent-space GG to enable backpropagation-based optimization. Results on many publicly available real-world image sets and a synthetic image set show the benefits of our method over existing methods.*

1. Introduction

Detecting abnormalities/anomalies/outliers is a well-known *one-class classification* (OCC) problem where learning relies mainly on data from the normal/inlier class. Here, completely characterizing the abnormal/outlier class using a finite-sized training sample is nearly infeasible because of the sample's inability to capture the large variability in the appearance of abnormalities. Consequently, such problems typically entail modeling the normal-class distribution, or its associated envelope as the decision boundary, in high-dimensional feature spaces. These distributions/boundaries are typically significantly non-Gaussian/nonlinear, thereby motivating the use of deep neural networks (DNNs) for effective modeling. Learning schemes for OCC typically train on data solely from the inlier class, but some recent OCC methods propose semi-supervised extensions that also leverage a small amount of training data from outlier classes. To deal with corrupted training sets, e.g., involving mislabeled or degraded images, some DNN methods use robust learning schemes.

We propose a novel *semi-supervised variational-learning* DNN framework leveraging *generalized-Gaussian* (GG) [22] models on both (i) encodings in latent space and (ii) the reconstructed image. We propose a generalized version of the variational autoencoder [15] (VAE) framework, namely, gVAE, leading to data-adaptive modeling subsuming robust modeling through the GG's shape parameters and uncertainty-aware modeling through the GG's scale parameters (as seen in Figure 1). We further extend that framework to a semi-supervised gVAE framework, namely, ss-gVAE, that can leverage some outlier data during training to improve performance. To enable backpropagation-based optimization, we propose a *reparameterization* of the GG. Results on many real-world image sets and a synthetic image set show the benefits of ss-gVAE over existing methods.

2. Related Work

In image analysis, some methods for OCC rely on variants of kernel principal component analysis (KPCA) [27, 13, 7, 16] or the support vector machine (SVM), i.e., the one-class SVM (OC-SVM) [26] and the support-vector data description (SVDD) [30]. Here, model learning relies on training data solely from the normal class, using hand-crafted image features and reproducing kernels. Other methods for OCC rely on density-based clustering (unsupervised and semi-supervised) of hand-crafted features based on the classic method by Hartigan [12], e.g., (i) DBSCAN: density based spatial clustering of applications with noise [9], (ii) its improved version DBSCAN* [3], and (iii) its hierarchical version HDBSCAN* [4].

Thanks to the Infrastructure Facility for Advanced Research and Education in Diagnostics grant funded by Department of Biotechnology, Government of India (BT/INF/22/SP23026/2017).

A class of methods rely on DNNs that automate optimal feature extraction and classifier learning by solving a single unified optimization problem. Some works [19, 33, 6, 34] on DNN-based OCC learn an autoencoder for the normal-class, such that the autoencoder will be unable to reconstruct (i.e., through a sequence of encoding followed by decoding) abnormal-class examples as accurately as normal-class examples. Methods like DRAE [33] typically outperform kernel-based methods like [17]. DeepSVDD [23] extended SVDD's strategy to DNNs that learn an encoder for normal-class images to map them to a compact subspace in the *latent* space, along with the assumption that the encoder will map abnormal-class images far from the aforementioned subspace. AnoGAN [25] uses a generative adversarial DNN to learn a manifold for normal data and their variability within the manifold. It uses a scoring scheme based on the mapping from image space to latent space, inferring anomalies based on their fit to the learned latent-space distribution. DRAE [33] and RCAE [6] enable learning from training sets corrupted with outliers by weeding out outliers or reducing their effects. RCAE [6] proposes an inductive learning scheme that extends robust PCA using a nonlinear autoencoder. Unlike all aforementioned methods, we combine variational learning and semi-supervised learning by extending the VAE framework [15, 32].

While OCC focuses mainly on learning from the normal class, a *small cohort of expert-labeled abnormal data* can help the classifier improve the estimation of the decision boundary enveloping the compact distribution of the latent-space encodings. While some non-DNN based methods [20, 1, 11] are able to leverage such limited information about the abnormal class using transductive learning to improve performance, typical DNN-based methods for OCC [19, 6, 23, 25] are unable to leverage such information. Among OCC methods, SSAD [11], QSSAE [29], and DeepSAD [24] leverage semi-supervision to learn a DNN-based one-class classifier. QSSAE [29] extends DCAE to leverage such limited supervision; analogous to RCAE, QSSAE aims to be robust to the corruption in the training data. The semi-supervised OCC methods of SSAD [11] and DeepSAD [24] relate to SVDD that focuses on mapping normal data to a compact subspace within latent space; while SSAD extends SVDD, DeepSAD extends DeepSVDD. While SSAD and SVDD rely on hand-crafted features, hand-crafted kernels, and unsupervised learning frameworks, DeepSAD relies on DNN-based semi-supervised learning that leverages a small training set of expert-labeled anomalous images.

3. Method

To classify images into normal and abnormal classes using DNN-based OCC, we extend the VAE framework [15] to a novel generalized VAE (gVAE) model (Section 3.1) and its learning formulation (Section 3.2). Our gVAE models both (i) the latent-space distribution conditioned on the input image (modeled by the encoder) as well as (ii) the output-image distribution conditioned on the latent-space encoding (modeled by the decoder) as GG distributions. The GG distributions lead to data-adaptive robust and uncertainty-aware modeling in both latent space and image space. To enable backpropagation within gVAE, Section 3.3 proposes a novel reparameterization (Equation 6) for the GG. To extend gVAE for semi-supervised learning, Section 3.4 proposes a novel ss-gVAE learning formulation (Equation 7) that improves performance by leveraging a small set of labeled outliers. Section 3.5 proposes a strategy for abnormality detection using ss-gVAE. Section 3.6 describes the DNN architecture and optimization strategy underlying ss-gVAE. Figure 1 illustrates gVAE and ss-gVAE.

3.1. A Generalized VAE (gVAE) Statistical Model

Let the random field X model an image in the space $\mathcal{X}$. Associated with the input image X, let Z model the *hidden/latent* random vector that captures all the information needed to generate the input image X. Let a DNN-based *encoder* model a mapping $\mathcal{E}(\cdot;\theta^{\mathcal{E}})$, parameterized by $\theta^{\mathcal{E}}$, which maps the input image X to the parameters $\mathcal{E}(X;\theta^{\mathcal{E}})$ of a distribution on the latent vector Z. Let a DNN-based *decoder* model a mapping $\mathcal{D}(\cdot;\theta^{\mathcal{D}})$, parameterized by $\theta^{\mathcal{D}}$, which maps the latent vector Z to the parameters $\mathcal{D}(Z;\theta^{\mathcal{D}})$ of a distribution on the image X. We propose to estimate the gVAE parameters $\theta^{\mathcal{ED}} := \theta^{\mathcal{E}} \cup \theta^{\mathcal{D}}$ by maximizing the likelihood of the observed training set images $\{X_n\}_{n=1}^{N}$.

Consider the joint probability density function (PDF) $P(X,Z)$ of the *complete data*, i.e., input image X and its latent-space encoding Z. Let $Q(Z|X;\theta^{\mathcal{E}}) \equiv Q(Z|\mathcal{E}(X;\theta^{\mathcal{E}}))$ be the conditional PDF of the latent-space encoding Z, conditioned on the input image X and parameterized by the encoder output $\mathcal{E}(X;\theta^{\mathcal{E}})$. Let $P(Z|X;\theta^{\mathcal{ED}})$ be the true latent-variable posterior PDF.

Let KL$(\cdot,\cdot)$ denote the Kullback-Leibler divergence between two PDFs. For each input image X, the log likelihood $\log P(X;\theta^{\mathcal{ED}}) =$

$$\mathcal{F}(Q(Z|X;\theta^{\mathcal{E}});\theta^{\mathcal{ED}}) + \text{KL}(Q(Z|X;\theta^{\mathcal{E}})\|P(Z|X;\theta^{\mathcal{ED}})), \tag{1}$$

where the functional $\mathcal{F}(Q(Z|X;\theta^{\mathcal{E}});\theta^{\mathcal{ED}})$ involves an expectation of the complete-data log-likelihood as $\mathcal{F}(Q(Z|X;\theta^{\mathcal{E}});\theta^{\mathcal{ED}}) := E_{Q(Z|X;\theta^{\mathcal{E}})}[\log P(X,Z;\theta^{\mathcal{ED}})] - E_{Q(Z|X;\theta^{\mathcal{E}})}[\log Q(Z|X;\theta^{\mathcal{E}})]$. Because KL divergence is non-negative, the functional $\mathcal{F}(\cdot)$ lower bounds the log-likelihood $\log P(X;\theta^{\mathcal{ED}})$. Let $P(Z)$ be a prior PDF on the latent random variable Z. Then, $\mathcal{F}(Q(Z|X;\theta^{\mathcal{E}});\theta^{\mathcal{ED}}) =$

$$E_{Q(Z|X;\theta^{\mathcal{E}})}[\log P(X|Z;\theta^{\mathcal{ED}})] - \text{KL}(Q(Z|X;\theta^{\mathcal{E}})\|P(Z)). \tag{2}$$

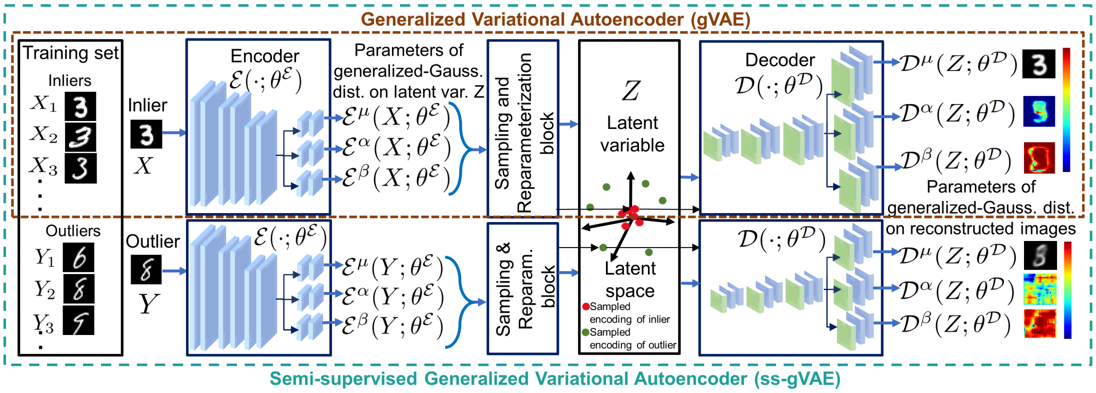

Generalized Variational Autoencoder (gVAE)
Training set
Inliers
X_1
X_2
X_3
Inlier
X
Encoder
$\mathcal{E}(\cdot;\theta^{\mathcal{E}})$
Parameters of generalized-Gauss. dist. on latent var. Z
$\mathcal{E}^{\mu}(X;\theta^{\mathcal{E}})$
$\mathcal{E}^{\alpha}(X;\theta^{\mathcal{E}})$
$\mathcal{E}^{\beta}(X;\theta^{\mathcal{E}})$
Sampling and Reparameterization block
Z
Latent variable
Decoder
$\mathcal{D}(\cdot;\theta^{\mathcal{D}})$
$\mathcal{D}^{\mu}(Z;\theta^{\mathcal{D}})$
$\mathcal{D}^{\alpha}(Z;\theta^{\mathcal{D}})$
$\mathcal{D}^{\beta}(Z;\theta^{\mathcal{D}})$
Parameters of generalized-Gauss. dist. on reconstructed images
Outliers
Y_1
Y_2
Y_3
Outlier
Y
$\mathcal{E}(\cdot;\theta^{\mathcal{E}})$
$\mathcal{E}^{\mu}(Y;\theta^{\mathcal{E}})$
$\mathcal{E}^{\alpha}(Y;\theta^{\mathcal{E}})$
$\mathcal{E}^{\beta}(Y;\theta^{\mathcal{E}})$
Sampling & Reparam. block
Latent space
● Sampled encoding of inlier
● Sampled encoding of outlier
$\mathcal{D}(\cdot;\theta^{\mathcal{D}})$
$\mathcal{D}^{\mu}(Z;\theta^{\mathcal{D}})$
$\mathcal{D}^{\alpha}(Z;\theta^{\mathcal{D}})$
$\mathcal{D}^{\beta}(Z;\theta^{\mathcal{D}})$
Semi-supervised Generalized Variational Autoencoder (ss-gVAE)

Figure 1. **Semi-Supervised Generalized VAE (ss-gVAE) Learning Framework for OCC.** For an input image, e.g., an inlier X, the DNN-encoder $\mathcal{E}(\cdot;\theta^{\mathcal{E}})$ outputs a factored GG PDF [22] on the multi-dimensional latent-space encoding Z, where $\theta^{\mathcal{E}}$ are the DNN-encoder parameters. The GG PDF is parameterized by a mean parameter vector $\mathcal{E}^{\mu}(\cdot;\theta^{\mathcal{E}})$, a log-scale parameter vector $\mathcal{E}^{\alpha}(\cdot;\theta^{\mathcal{E}})$, and a log-shape parameter vector $\mathcal{E}^{\beta}(\cdot;\theta^{\mathcal{E}})$. For a specific latent-space encoding Z, the DNN-decoder $\mathcal{D}(\cdot;\theta^{\mathcal{D}})$ outputs a factored GG PDF in image space, where $\theta^{\mathcal{D}}$ are the DNN-encoder parameters. Variational inference entails sampling in latent space, and our novel scheme for reparameterization of the samples from the latent-space GG enables backpropagation-based optimization, thereby significantly extending the VAE framework [15]. Our introduction of the GG enables data-adaptive modeling subsuming both robust statistical modeling and uncertainty-aware modeling; this novel framework is **gVAE**. We further extend gVAE for semi-supervised learning to leverage a small number of labeled outliers, i.e., input images Y, promoting separation of the distributions of encodings of inliers and outliers in latent space; this novel framework is **ss-gVAE**. In this example using MNIST data, we treat images of digit 3 as inliers and images of all other digits as outliers. Given a test image U, we classify it as inlier/outlier based on the norm of the encoder-mapped mean vector $\mathcal{E}^{\mu}(U;\theta^{\mathcal{E}})$.

Modeling the True Posterior. The true posterior $P(Z|X;\theta^{\mathcal{ED}})$ is unknown. Even though the PDF $P(X)$ on input images can have a complex structure, a sufficiently non-linear encoder mapping can transform $P(X)$ to a tractable latent-space posterior PDF $P(Z|X;\theta^{\mathcal{ED}})$ that is a factored product of GG PDFs. The univariate GG [22] with mean μ, scale c, and shape ρ, is $G(u;\mu,c,\rho) :=$

$$\frac{\delta(\rho/2)}{2\sqrt{c}} \exp\left(-\left(\eta(\rho/2)\|u-\mu\|^2/c\right)^{\rho/2}\right), \qquad (3)$$

where $\delta(r) := r\Gamma(2/r)/(\pi\Gamma(1/r)^2)$, $\Gamma(\cdot)$ denotes the gamma function, and $\eta(r) := \Gamma(2/r)/(2\Gamma(1/r))$.

Designing the Prior $P(Z)$. We extend the motivation for modeling the true-posterior as a product of GG PDFs to motivate, without loss of generality, that the factored GG $P(Z|X;\theta^{\mathcal{ED}})$ can be close to a standard multivariate normal. So, we model the prior PDF $P(Z)$ as a standard multivariate normal.

Designing Encoder Output and Latent-Space PDF. If KL $(Q(Z|X;\theta^{\mathcal{E}})\|P(Z|X;\theta^{\mathcal{ED}}))$ is close to zero, then the log-likelihood $\log P(X;\theta^{\mathcal{ED}})$ is close to the functional $\mathcal{F}(Q(Z|X;\theta^{\mathcal{E}});\theta^{\mathcal{ED}})$, and we can learn the DNN parameters $\theta^{\mathcal{ED}}$ by optimizing the functional instead of the log-likelihood. So, we propose to model $Q(Z|X;\theta^{\mathcal{E}})$ by the same (factored) form motivated for the posterior PDF $P(Z|X;\theta^{\mathcal{ED}})$. Then, for each input X, the encoder output parameterizes the corresponding multivariate (factored) GG $Q(Z|X;\theta^{\mathcal{E}})$ using (i) a mean vector $\mathcal{E}^{\mu}(X;\theta^{\mathcal{E}})$, (ii) a vector of log-scale parameters $\mathcal{E}^{\alpha}(X;\theta^{\mathcal{E}})$, and (iii) a vector of log-shape parameters $\mathcal{E}^{\beta}(X;\theta^{\mathcal{E}})$.

Designing Decoder Output and Loss. We design the decoder output $\mathcal{D}(Z;\theta^{\mathcal{D}})$ to model a distribution on the input image X, which allows for heavy-tailed residuals and models heteroscedasticity across pixels. Specifically, we design the decoder output $\mathcal{D}(Z;\theta^{\mathcal{D}})$ to be the parameters of a factored GG PDF $P(X|Z)$. Thus, the decoder output parameterizes the multivariate (factored) GG $P(X|Z;\theta^{\mathcal{D}})$ using (i) a mean vector $\mathcal{D}^{\mu}(Z;\theta^{\mathcal{D}})$, (ii) a vector of log-scale parameters $\mathcal{D}^{\alpha}(Z;\theta^{\mathcal{D}})$, and (iii) a vector of log-shape parameters $\mathcal{D}^{\beta}(Z;\theta^{\mathcal{D}})$. Consequently, the loss function equals the negative log-likelihood $-\log P(X|Z;\theta^{\mathcal{D}})$.

3.2. Unsupervised Learning with gVAE

Let $\{X_n\}_{n=1}^{N}$ model a training set representing data from the normal class. Analogous to the motivation underlying the learning formulation of the VAE [15], we propose to learn the parameters $\theta^{\mathcal{ED}}$ underlying our gVAE by minimizing the expected loss for all the (normal) data points as

$$\arg\min_{\theta^{\mathcal{ED}}} \mathcal{L}_{\text{normal}}(X;\theta^{\mathcal{ED}}), \text{where}$$

$$\mathcal{L}_{\text{normal}}(X;\theta^{\mathcal{ED}}) := \frac{1}{NI}\sum_{n=1}^{N}\sum_{i=1}^{I}\Big[0.5\|z_{ni}(X_n,\theta^{\mathcal{E}})\|_2^2 + \log\mathcal{G}(z_{ni}(X_n,\theta^{\mathcal{E}});\mathcal{E}^{\mu}(X_n;\theta^{\mathcal{E}}),\mathcal{E}^{\alpha}(X_n;\theta^{\mathcal{E}}),\mathcal{E}^{\beta}(X_n;\theta^{\mathcal{E}})) - \log\mathcal{G}(X_n;\mathcal{D}^{\mu}(Z_n;\theta^{\mathcal{D}}),\mathcal{D}^{\alpha}(Z_n;\theta^{\mathcal{D}}),\mathcal{D}^{\beta}(Z_n;\theta^{\mathcal{D}}))\Big], \qquad (4)$$

where (i) $z_{ni}(X_n, \theta^{\mathcal{E}})$ represents the i-th independent draw from the PDF $Q(Z|X_n;\theta^{\mathcal{E}}) = \mathcal{G}(Z; \mathcal{E}^{\mu}(X_n;\theta^{\mathcal{E}}), \mathcal{E}^{\alpha}(X_n;\theta^{\mathcal{E}}), \mathcal{E}^{\beta}(X_n;\theta^{\mathcal{E}}))$, where the draw $z_{ni}(X_n, \theta^{\mathcal{E}})$ is actually a function of the encoder parameters $\theta^{\mathcal{E}}$ and the n-th input X_n, and (ii) $\mathcal{G}(\cdot;\mu,\alpha,\beta)$ is the factorized GG with mean vector μ, a vector of log-scale parameters α, and a vector of log-shape parameters β. For numerical stability and differentiability, we propose to evaluate $\log \mathcal{G}(a;\mu,\alpha,\beta)$ as follows. Let the operator $[\cdot]_d$ denote the d-th scalar component of its vector argument. Then, $[\log \mathcal{G}(b;\mu,\alpha,\beta)]_d :=$

$$\begin{aligned}&\log((\tau + \exp([\beta]_d))/(\Delta + \exp([\alpha]_d))) \\ &- \left(\left(\frac{b - [\mu]_d}{\Delta + \exp([\alpha]_d)} \right)^2 + \epsilon \right)^{0.5(\tau + \exp([\beta]_d))} \\ &- \log(\Gamma(1/(\tau + \exp([\beta]_d))) + \text{constant}, \qquad (5)\end{aligned}$$

where Δ, ϵ, and τ are small positive real-valued constants that regularize the function to ensure differentiability.

3.3. Reparameterizing the Generalized-Gaussian

To enable backpropagation for $\theta^{\mathcal{E}}$, we propose to reparameterize the i-th independent draw $z_{ni} \sim \mathcal{G}(Z; \mathcal{E}^{\mu}(X_n;\theta^{\mathcal{E}}), \mathcal{E}^{\alpha}(X_n;\theta^{\mathcal{E}}), \mathcal{E}^{\beta}(X_n;\theta^{\mathcal{E}}))$ using a hierarchical reparameterization scheme. Let Gamma(a, b) be the Gamma PDF with shape parameter a and scale parameter b. We first reparameterize [21] the GG random variable based on a Gamma random variable as

$$\begin{aligned}&[z_{ni}]_d := [\mathcal{E}^{\mu}(X_n;\theta^{\mathcal{E}})]_d + \\ &(\Delta + [\exp(\mathcal{E}^{\alpha}(X_n;\theta^{\mathcal{E}})]_d) S_{nid} |Y_{nid}|^{1/(\tau + \exp([\mathcal{E}^{\beta}(X_n;\theta^{\mathcal{E}})]_d))}, \qquad (6)\end{aligned}$$

where (i) random variable S_{nid} takes values $+1$ or -1 with equal probability and (ii) random variable Y_{nid} has the PDF Gamma$(1/(\tau + [\mathcal{E}^{\beta}(X;\theta^{\mathcal{E}})]_d), 1)$. Subsequently, we leverage implicit reparameterization gradients [10] to enable backpropagation across the Gamma random variable.

3.4. Semi-supervised gVAE (ss-gVAE) Learning

In addition to the training set $\{X_n\}_{n=1}^{N}$ representing data from the normal class, our ss-gVAE framework leverages a *much smaller* set of expert-labeled outliers, say, $\{Y_m\}_{m=1}^{M}$, where typically $M \ll N$, to improve the learning over our gVAE framework. While the gVAE objective function in (Equation 4) seeks high values of the log-likelihood $\log P(X;\theta^{\mathcal{E}\mathcal{D}})$, the ss-gVAE objective function includes an additional term that simultaneously seeks low values of the log-likelihood $\log P(Y;\theta^{\mathcal{E}\mathcal{D}})$ of the outliers. Let $z_{mj}(Y_m, \theta^{\mathcal{E}})$ represent the j-th independent draw from the PDF $Q(Z|Y_m;\theta^{\mathcal{E}}) = \mathcal{G}(Z; \mathcal{E}^{\mu}(Y_m;\theta^{\mathcal{E}}), \mathcal{E}^{\alpha}(Y_m;\theta^{\mathcal{E}}), \mathcal{E}^{\beta}(Y_m;\theta^{\mathcal{E}}))$, which is actually a function of the encoder parameters $\theta^{\mathcal{E}}$ and the m-th outlier input Y_m. Let $z_{mj}(Y_m, \theta^{\mathcal{E}})$ also be reparameterized using the same strategy as proposed earlier for $z_{ni}(X_n, \theta^{\mathcal{E}})$.

We formulate ss-gVAE learning as

$$\arg\min_{\theta^{\mathcal{E}\mathcal{D}}} \eta \mathcal{L}_{\text{normal}}(X;\theta^{\mathcal{E}\mathcal{D}}) - (1-\eta)\mathcal{L}_{\text{abnormal}}(Y;\theta^{\mathcal{E}\mathcal{D}}), \qquad (7)$$

where

$$\begin{aligned}\mathcal{L}_{\text{abnormal}}(Y;\theta^{\mathcal{E}\mathcal{D}}) := \frac{1}{MJ}\sum_{m=1}^{M}\sum_{j=1}^{J}\Big[0.5\|z_{mj}(Y_m, \theta^{\mathcal{E}})\|_2^2 + \\ \log \mathcal{G}(z_{mj}(Y_m, \theta^{\mathcal{E}}); \mathcal{E}^{\mu}(Y_m;\theta^{\mathcal{E}}), \mathcal{E}^{\alpha}(Y_m;\theta^{\mathcal{E}}), \mathcal{E}^{\beta}(Y_m;\theta^{\mathcal{E}})) \\ - \log \mathcal{G}(Y_m; \mathcal{D}^{\mu}(Z_m;\theta^{\mathcal{D}}), \mathcal{D}^{\alpha}(Z_m;\theta^{\mathcal{D}}), \mathcal{D}^{\beta}(Z_m;\theta^{\mathcal{D}}))\Big], \qquad (8)\end{aligned}$$

where free parameter $\eta \in (0, 1)$ weights the two parts of the objective function, i.e., one relying on the inlier set and another relying on the outlier set. We tune η relying on a small validation set comprising inliers and outliers.

3.5. Inference Strategy for Abnormality Detection

The standard-normal prior $P(Z)$ associated with the distribution of the latent-space encodings of the normal-class data X promotes the latent-space distribution $Q(Z|X;\theta^{\mathcal{E}})$ to be close to $P(Z)$ and, thereby, promotes the encoded mean vector $\mathcal{E}^{\mu}(X;\theta^{\mathcal{E}})$ to be close to the origin. Similarly, for normal data X, the decoder-based loss $-\log P(X|Z;\theta^{\mathcal{D}})$ promotes the decoder-output mean vector $\mathcal{D}^{\mu}(Z;\theta^{\mathcal{D}})$ to be a good reconstruction of (i.e., be close to) the input X. During semi-supervised learning, for the outlier data Y, the objective function promotes the encoded mean vector $\mathcal{E}^{\mu}(Y;\theta^{\mathcal{E}})$ to be away from the origin and the decoder-output mean vector $\mathcal{D}^{\mu}(Z;\theta^{\mathcal{D}})$ to be far from the input Y. Thus, for a test image U that needs to be classified, in principle, an inference strategy can rely on the latent-space encoding, e.g., through $\mathcal{E}^{\mu}(U;\theta^{\mathcal{E}})$, or the decoded output, e.g., through $\mathcal{D}^{\mu}(Z;\theta^{\mathcal{D}})$, or both. For the datasets in this paper, we find that relying solely in the latent-space encoding suffices. Thus, we propose an inference strategy that first computes the anomaly score as the norm of the mean vector $\mathcal{E}^{\mu}(U;\theta^{\mathcal{E}})$ associated with the latent-space PDF for U (a larger score makes it more likely for U to be an outlier), and then use a threshold ρ on the score to classify U as inlier/outlier. ρ is a free parameter that we tune using a small validation set comprising inliers and outliers.

3.6. DNN Architecture and Optimization Strategy

Figure 1 illustrates the DNN architecture. The encoder $\mathcal{E}(\cdot;\theta^{\mathcal{E}})$ comprises a common block before forking into three branches that output the mean $\mathcal{E}^{\mu}(X;\theta^{\mathcal{E}})$, log-scale parameters $\mathcal{E}^{\alpha}(X;\theta^{\mathcal{E}})$, and log-shape parameters $\mathcal{E}^{\beta}(X;\theta^{\mathcal{E}})$. The common block comprises three sub-blocks, where each sub-block models convolutions, pooling, batch normalization, and leaky-ReLU activation. Subsequently, each branch has one sub-block modeling a fully-connected layer and leaky-ReLU activation. The decoder is analogous to the encoder, comprising (i) a common block with three sub-blocks, where each sub-block models upsampling, convolutions, batch normalization, and leaky-ReLU activation, and (ii) three branches that output the mean $\mathcal{D}^{\mu}(Z;\theta^{\mathcal{D}})$, log-scale parameters $\mathcal{D}^{\alpha}(Z;\theta^{\mathcal{D}})$, and log-shape parameters $\mathcal{D}^{\beta}(Z;\theta^{\mathcal{D}})$, where each branch has one sub-block as in (i).

Our training strategy is sequential. First, we use solely the inlier training set and exclude the branches that model log-scale and log-shape (fixing GG scale parameters to 1 and shape parameters to 2), thereby training the DNN as a VAE. Second, with this warm start, we include the log-scale branch and the log-shape branch and optimize the parameters underlying those branches, thereby training the DNN as a gVAE. Finally, we include the outlier set as well, and retrain the entire DNN, thereby training the DNN as a ss-gVAE. We tune the free parameters using a small validation set comprising inliers and outliers.

4. Results and Discussion

We evaluate on several real-world datasets (MNIST, Fashion-MNIST, CIFAR-10, 10 MVTec datasets, Malaria dataset) and a synthetic image set.

We compare with 6 *baseline methods*: (i) DeepSAD [24]: semi-supervised DNN-based OCC using both $\{X_n\}_{n=1}^{N}$ and $\{Y_m\}_{m=1}^{M}$ for training; (ii) SSAD-Hybrid [5]: semi-supervised kernel-based OCC extending SSAD [11] using pre-extracted autoencoder-based features; (iii) DeepSVDD [23]: unsupervised DNN-based OCC using only $\{X_n\}_{n=1}^{N}$ for training; (iv) OC-SVM-Hybrid [8, 5]: unsupervised kernel-based OCC extending OC-SVM [28] using pre-extracted autoencoder-based features; (v) ss-DCAE: semi-supervised version of DNN-based OCC [19] relying on image reconstruction using an autoencoding strategy; and (vi) BinClass: fully-supervised DNN-based binary classifier using both inliers and outliers for training. We evaluate each method at different levels of supervision $\gamma := M/(M+N)$ ranging within $[0, 0.2]$; indeed, BinClass cannot perform at level of supervision $\gamma = 0$. In this paper, all kernel-based methods use the radial-basis-function kernel. We split the available data into 3 mutually-exclusive and exhaustive subsets for training, validation (to tune the free parameters underlying each method), and testing. For quantitative evaluation, we use the area under the receiver-operating-characteristics curve (AUC).

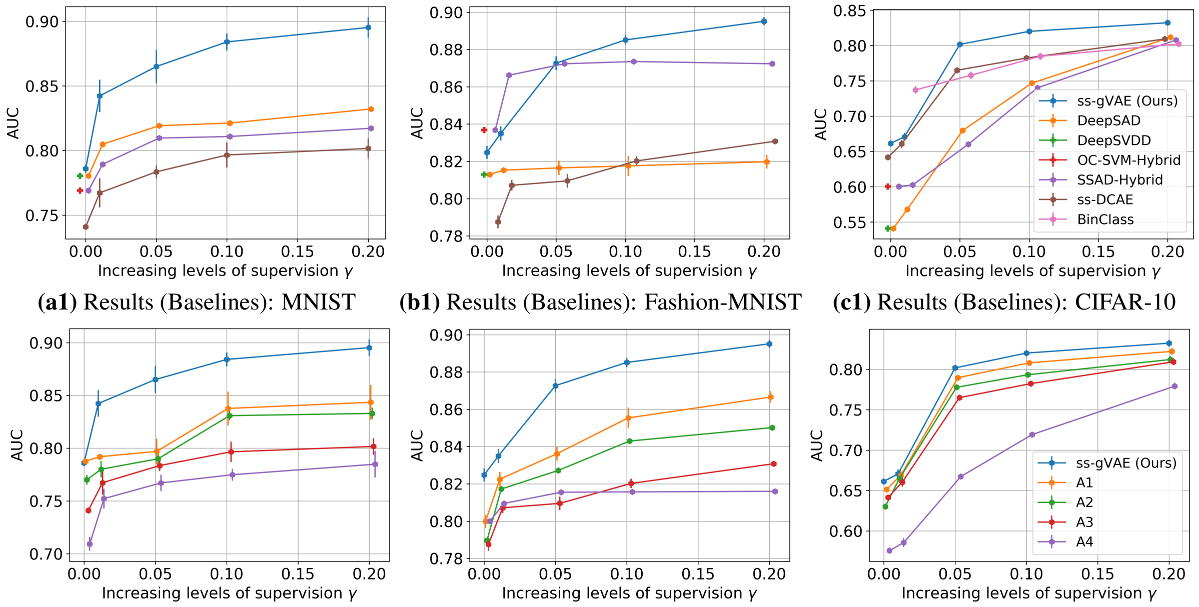
We also perform *ablation studies* to gain insights into the various components of our method. We denote various ablated versions of our method ss-gVAE as follows. (i) **A1**: removes from ss-gVAE the latent-space components of GG modeling and variational learning, but includes a latent-space loss of the squared distance of the encoding from the origin (similar to the idea underlying SVDD-based methods); (ii) **A2**: removes from A1 the GG modeling components in image space, thereby reducing the decoder-based loss by the squared norm of the reconstruction residual; (iii) **A3**: removes from A2 the squared-norm based loss in latent space, thereby making A3 akin to semi-supervised DCAE; (iv) **A4**: removes from A2 the decoder-based loss term in image space, thereby making A4 akin to DeepSAD.

4.1. Results on MNIST, Fashion-MNIST, CIFAR-10

For each of these datasets, we consider the images in 1 of the classes as inlier/normal, and the images in all other 9 classes as outlier/abnormal; we repeat experiments by choosing each of the 10 classes as inlier (and the rest as outliers), and then pool the results. During evaluation, the test set comprises images from all 10 classes (equal number of inliers and outliers), to be classified into inlier or outlier. During semi-supervised learning, to create the expert-labeled training set comprising outliers, we use images from only 1 of the other 9 classes; this mimics the real-world scenario that motivates the OCC strategy itself.

For all DNN methods, the architecture uses a series of 3 blocks, each comprising convolutional, batch normalization, and leaky-ReLU layers. After the last convolutional block in the encoder, we use 3 fully-connected layers, in parallel, to get the latent embeddings for the mean, log-scale, and log-shape. The latent-space dimension for MNIST, Fashion-MNIST, and CIFAR-10 is 32, 64, and 128, respectively. The decoder complements the encoder, typical for an autoencoder. We use Adam [14]; batch size 128; weight decay $\lambda = 5 \times 10^{-7}$. Following the evaluation strategy of SSAD in [11] and DeepSAD in [23], to evaluate the robustness of the learning to imperfectly curated data, we introduce pollution/corruption in the training set in the form of 20% of the outlier images being misclassified as inliers.

The results (Figure 2(a1)–(c1)) show that, compared to other methods, our method (ss-gVAE) performs better than all other methods at virtually every level of supervision. At non-zero levels of semi-supervision, SSAD-Hybrid is able to improve upon its unsupervised version, i.e., OC-SVM-Hybrid. Similarly, DeepSAD improves over its unsupervised version, i.e., DeepSVDD. While one category of baselines, i.e., OC-SVM-Hybrid, SSAD-Hybrid, DeepSVDD, and DeepSAD, incorporate only the encoder representations during learning, another category of methods, e.g., ss-DCAE, incorporates the autoencoded/reconstructed images for model learning. While performance on some datasets (i.e., MNIST, Fashion-MNIST) is better when models use the former strategy, performance on other datasets (i.e., CIFAR-10) is better for autoencoder/reconstruction based models. This is indeed the motivation for gVAE/ss-gVAE



(a1) Results (Baselines): MNIST **(b1)** Results (Baselines): Fashion-MNIST **(c1)** Results (Baselines): CIFAR-10

(a2) Results (Ablation): MNIST **(b2)** Results (Ablation): Fashion-MNIST **(c2)** Results (Ablation): CIFAR-10

Figure 2. **Results on MNIST, Fashion-MNIST, CIFAR-10.** Across different levels of supervision γ, AUC values for: **(a1)–(c1) 6 baselines** (BinClass performs very poorly for (a1)-(b1) and is absent from those two plots) and **(a2)–(c2) 4 ablated versions** of ss-gVAE. The plots (with bars) indicate the variability in AUC across randomly sampled training sets, validation sets, and test sets (20 repeats).

that learn distributions in latent space as well as reconstructions in image space. Moreover, our GG models incorporate, both, robust heavy-tailed modeling (through the shape parameters) and uncertainty-aware heteroscedastic modeling (through the scale parameters).

Figure 2(a2)–(c2) shows that the ablated versions relying either solely on the reconstruction-based losses (i.e., A3 that is akin to semi-supervised DCAE) or solely on the latent-space losses (i.e., A4, equivalent to DeepSAD) perform poorer than A2 that includes a combination of both latent-space loss and image-space losses. A1, i.e., the ablated version of ss-gVAE excluding variational learning, is unable to perform as good as our proposed approach of ss-gVAE. Moreover, the benefits of our GG modeling are demonstrated by the improved performance of ss-gVAE and A1 over their ablated versions (i.e., A2, A3, and A4) that remove all GG components.

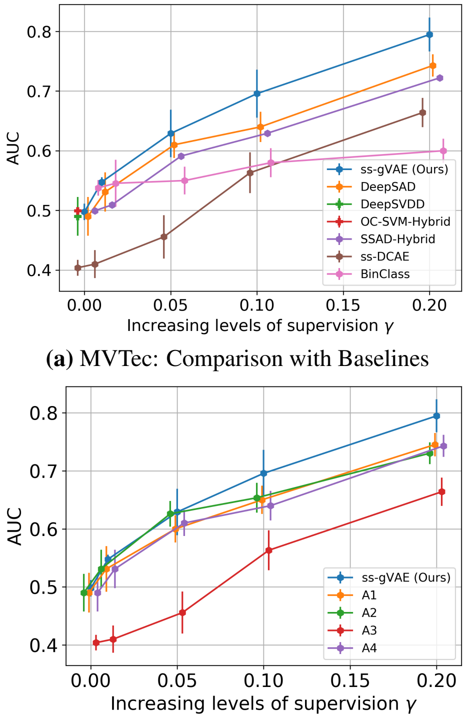
4.2. Results on 10 MVTec Datasets

The MVTec dataset [2] is specifically designed for OCC, including a spectrum of categories of textures and objects, which are each subdivided further into 3–8 subcategories. As motivated earlier, we introduce pollution/corruption in the training set in the form of 20% of outlier images being misclassified as inliers. For the texture categories, because the abnormality is present in a tiny part of the entire image, we learn all models on patches of size 64 × 64 pixels with and without containing the abnormality. During semi-supervised learning, to create the expert-labeled training set comprising outliers, we use images from about half the subcategories. During evaluation, the test set comprises images from all the subcategories to be classified into inlier/normal or outlier/abnormal. We resize the patches to 32×32 pixels and use the same architecture as for CIFAR-10.

Figure 3 and Table 1 show trends similar to those in the previous subsection. ss-gVAE improves over other baselines at virtually every level of supervision. Here as well, at non-zero levels of semi-supervision, SSAD-Hybrid and DeepSAD improve over their unsupervised versions, i.e., OC-SVM-Hybrid and DeepSVDD, respectively. On about half the categories, ss-DCAE (reconstruction-based scheme) performs better than SSAD-Hybrid (latent-space based scheme), while the behaviour is reverse on the other categories. The improvement of ss-gVAE over other baselines (Figure 3(a)) stem from ss-gVAE's modeling and learning of GG distributions in both latent space and image space. ss-gVAE's improvements over the ablated versions A1 and A2 (Figure 3(b)) stem from the variational inference coupled with GG modeling incorporating the principles of robust heavy-tailed modeling (through the shape parameter) and uncertainty-aware heteroscedastic modeling (through the scale parameter). All semi-supervised methods show significant improvement even with 5% supervision ($\gamma = 0.05$) over their counterparts that cannot leverage semi-supervision (e.g., DeepSVDD, OC-SVM-Hybrid, ss-DCAE). BinClass cannot operate at $\gamma = 0$. For larger γ, BinClass does perform better than some methods, but typically poorer than DeepSAD or ss-gVAE; indeed, BinClass

Table 1. **Results on 10 MVTec Datasets (5 Textures, 5 Objects): Comparing 6 Baselines.** AUC mean (± standard deviation) for each method shows the variability in performance across randomly sampled training sets, validation sets, and test sets (20 repeats).

Category	**ss-gVAE**	DeepSAD	DeepSVDD	SSAD-Hybrid	OC-SVM-Hybrid	ss-DCAE	BinClass
	$\gamma = 0.2$	$\gamma = 0.2$	unsupervised	$\gamma = 0.2$	unsupervised	$\gamma = 0.2$	$\gamma = 0.2$
Texture: Carpet	$\mathbf{79.5 \pm 2.8}$	74.3 ± 1.8	49.0 ± 3.2	72.23 ± 0.2	49.9 ± 0.1	66.4 ± 2.4	60.0 ± 0.2
Texture: Grid	71.9 ± 2.4	72.5 ± 2.7	69.1 ± 1.9	66.86 ± 0.4	63.8 ± 0.1	59.0 ± 7.9	$\mathbf{76.3 \pm 1.2}$
Texture: Leather	92.2 ± 0.3	93.7 ± 0.8	85.7 ± 0.4	$\mathbf{96.4 \pm 0.3}$	94.1 ± 0.2	82.9 ± 1.9	89.4 ± 0.1
Texture: Tile	$\mathbf{76.3 \pm 1.4}$	63.6 ± 2.2	53.8 ± 4.1	74.3 ± 0.1	52.1 ± 0.2	75.4 ± 1.0	55.8 ± 2.5
Texture: Wood	$\mathbf{76.3 \pm 3.3}$	76.2 ± 2.5	74.1 ± 2.9	67.2 ± 0.1	54.0 ± 0.7	73.0 ± 1.5	75.1 ± 1.5
Textures: Average	$\mathbf{79.2 \pm 2.0}$	76.1 ± 2.0	66.3 ± 2.5	75.4 ± 0.2	62.8 ± 0.3	71.3 ± 2.9	71.3 ± 1.1
Object: Bottle	$\mathbf{82.1 \pm 4.2}$	76.9 ± 2.3	68.2 ± 2.2	52.7 ± 2.0	50.0 ± 0.2	70.6 ± 5.2	76.2 ± 4.2
Object: Cable	$\mathbf{75.4 \pm 1.1}$	69.2 ± 1.1	61.1 ± 2.2	57.8 ± 1.1	55.1 ± 0.8	61.7 ± 0.6	68.9 ± 3.2
Object: Hazelnut	$\mathbf{60.0 \pm 1.7}$	57.2 ± 2.1	54.8 ± 0.6	53.0 ± 0.6	51.4 ± 0.4	53.5 ± 2.2	50.3 ± 0.4
Object: Metal nut	$\mathbf{75.5 \pm 3.2}$	71.5 ± 1.5	69.1 ± 0.8	59.2 ± 0.3	58.0 ± 0.2	45.0 ± 2.3	74.0 ± 0.2
Object: Screw	$\mathbf{62.3 \pm 0.8}$	61.2 ± 2.5	53.1 ± 1.8	61.9 ± 0.7	58.2 ± 0.4	58.0 ± 1.7	59.3 ± 1.5
Objects: Average	$\mathbf{71.1 \pm 2.2}$	67.2 ± 1.9	61.3 ± 1.5	56.9 ± 0.9	54.5 ± 0.4	57.8 ± 2.4	65.7 ± 1.9



(a) MVTec: Comparison with Baselines

(b) MVTec: Ablation Studies

Figure 3. **Results on MVTec Carpet Category.** AUC values for: **(a)** baseline methods and **(b)** ablated versions of ss-gVAE. The plots (with bars) indicate the variability in AUC across randomly sampled training sets, validation sets, and test sets (20 repeats).

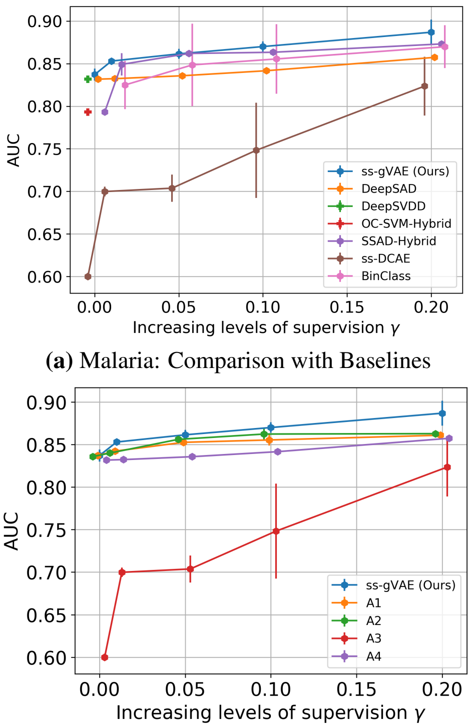
is prone to overfitting and poor generalization, because the training set is unable to capture all the variability of the abnormalities present in the test set (thereby motivating OCC).

4.3. Results on a Dataset of Malaria-Infected Cells

The Broad Bioimage Benchmark Collection [18] comprises 1328 images. Each image contains red blood cells (RBCs) that are labeled as non-infected (normal) and infected (abnormal); RBC diameter here is typically 170 pixels. So, we aim to classify patches of size 170×170 pixels; we consider a patch as abnormal if at least 50% of its pixels are labeled abnormal. We extract 10,000 normal RBCs and 2,500 abnormal RBCs. The infected RBCs are of 6 types. During training, we use only 3 types of abnormal RBCs. During testing, we use abnormal RBCs of all types. The test set contains an equal number of normal and abnormal RBCs. We introduce pollution/corruption in the training set in the form of 10% of outlier images being misclassified as inliers. We resize the patches to 32×32 pixels and use the same architecture as for CIFAR-10. ss-gVAE outperforms other baselines at virtually all levels of supervision (Figure 4). The t-SNE [31] plots (Figure 5) show improved separability between the normal and abnormal data for ss-gVAE compared to other baselines. Moreover, ss-gVAE depicts fewer misclassification errors. These qualitative analyses are consistent with the quantitative analyses (Figure 4). With increasing supervision, the separability of the normal and outlier classes, as well as the classification accuracy, increases. In both these aspects, ss-gVAE improves significantly over DeepSAD.

4.4. Results on a Synthetic Dataset

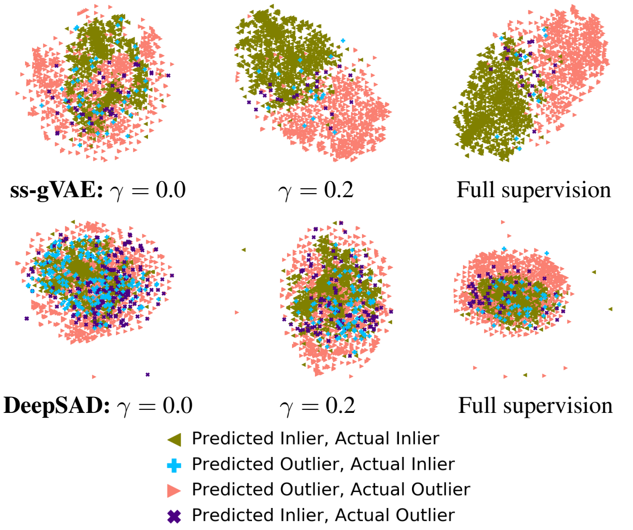
We designed a synthetic image set with 1250 RGB images of size 32×32 pixels. For the normal class, the intensities in each color channel were drawn independently from Gaussian distributions with the mean parameters chosen randomly (uniformly) within $[0, 1]$ and the scale parameters chosen randomly (uniformly) within $[0.5, 1.5]$. For the abnormal class, the intensities in each color channel were drawn independently from Gaussians with the mean parameters chosen in the same way as for the normal class, and the scale parameters chosen randomly (uniformly) within $[1.5, 2.5]$. Thus, the differentiation between the classes relies on the subtle higher-order textural features within the underlying images. The qualitative analysis in Figure 7 and



(a) Malaria: Comparison with Baselines

(b) Malaria: Ablation Studies

Figure 4. **Results on Malaria Dataset.** AUC values for: **(a)** baseline methods and **(b)** ablated versions of ss-gVAE. The plots (with bars) indicate the variability in AUC across randomly sampled training sets, validation sets, and test sets (20 repeats).



ss-gVAE: $\gamma = 0.0$ $\gamma = 0.2$ Full supervision

DeepSAD: $\gamma = 0.0$ $\gamma = 0.2$ Full supervision

Figure 5. **Results on Malaria Dataset: t-SNE visualizations of latent-space distributions** from the test set for normal and abnormal RBCs with increasing level of supervision γ.

quantitative analysis in Figure 6, both, show the improvements of ss-gVAE over DeepSAD and other baselines, in ways similar to those seen in earlier sections.

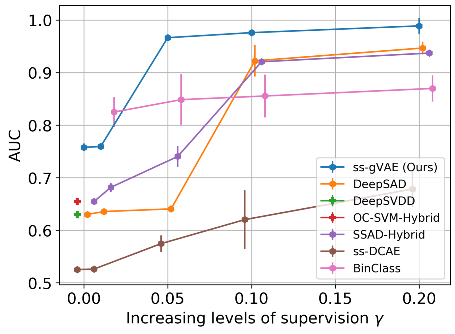
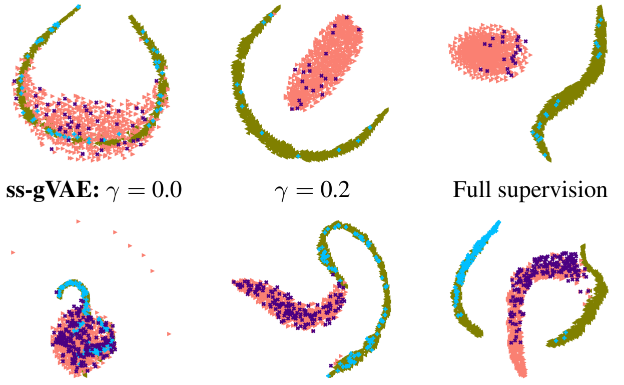


Figure 6. **Results on a Synthetic Dataset.** AUC values for ss-gVAE and the baseline methods. The plots (with bars) indicate the variability in AUC across randomly sampled training sets, validation sets, and test sets (20 repeats).



ss-gVAE: $\gamma = 0.0$ $\gamma = 0.2$ Full supervision

DeepSAD: $\gamma = 0.0$ $\gamma = 0.2$ Full supervision

Figure 7. **Results on a Synthetic Dataset: t-SNE visualizations of latent-space distributions** for normal and abnormal test data at varying levels of supervision γ. [Graph legend same as that in Figure 5]

5. Conclusion

This paper proposes a novel generalization of the VAE framework for OCC. Our novel gVAE framework leverages factored GG models that incorporate, both, robust heavy-tailed modeling (through a set of shape parameters) and uncertainty-aware heteroscedastic modeling (through a set of scale parameters) in image space. The shape parameters and scale parameters are learned and output by the gVAE separately for each input datum, thereby leading to data-adaptive modeling. We propose a novel reparameterization for sampling from the latent-space GG to enable backpropagation-based optimization. We further extend gVAE to a novel ss-gVAE framework that improves performance by leveraging a small set of labeled outliers in the training set. Comprehensive empirical analyses, involving 6 baselines and 4 ablated versions on several publicly available datasets and a synthetic image set, show the benefits of our method over the state of the art and provide insights into the benefits of the novel components.

References

[1] E Bauman and K Bauman. One-class semi-supervised learning. In *Braverman Readings in Machine Learning. Key Ideas from Inception to Current State*, pages 189–200, 2017.

[2] P Bergmann, M Fauser, D Sattlegger, and C Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *IEEE Comp. Vis. Pattern Recog.*, pages 9592–600, 2019.

[3] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*, pages 160–172. Springer, 2013.

[4] Ricardo JGB Campello, Davoud Moulavi, Arthur Zimek, and Jörg Sander. Hierarchical density estimates for data clustering, visualization, and outlier detection. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 10(1):1–51, 2015.

[5] V L Cao, M Nicolau, and J McDermott. A hybrid autoencoder and density estimation model for anomaly detection. In *Parallel Problem Solving from Nature*, pages 717–726, 2016.

[6] R Chalapathy, AK Menon, and S Chawla. Robust, deep and inductive anomaly detection. In *Eur. Conf. Mach. Learn. Know. and Princ. Pract. Knowl. Disc. in Data.*, pages 36–51, 2017.

[7] R Das, A Golatkar, and SP Awate. Sparse kernel pca for outlier detection. In *IEEE Int. Conf. Machine Learning and Appl.*, pages 152–7, 2018.

[8] S M Erfani, S Rajasegarar, S Karunasekera, and C Leckie. High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning. *Pattern Recognition*, 58:121 – 134, 2016.

[9] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise. *KDD*, 96(34):226–31, 1996.

[10] M Figurnov, S Mohamed, and A Mnih. Implicit reparameterization gradients. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31, pages 441–452. Curran Associates, Inc., 2018.

[11] N Gornitz, M Kloft, K Rieck, and U Brefeld. Toward supervised anomaly detection. *J. Artif. Int. Res.*, 46(1):235–262, 2013.

[12] John A Hartigan. *Clustering algorithms*. John Wiley & Sons, Inc., 1975.

[13] H Hoffmann. Kernel PCA for novelty detection. *Pattern Recog.*, 40(3):863–74, 2007.

[14] D Kingma and B Jimmy. Adam: A method for stochastic optimization. *Int. Conf. Learn. Rep.*, 2015.

[15] D Kingma and M Welling. Auto-encoding variational bayes. In *2nd Int Conf on Lear Rep, ICLR 2014*, 2014.

[16] Nitin Kumar and Suyash P Awate. Semi-supervised robust mixture models in rkhs for abnormality detection in medical images. *IEEE Trans. Image Proc.*, 29:4772–87, 2020.

[17] W Liu, G Hua, and J Smith. Unsupervised one-class learning for automatic outlier removal. In *IEEE Comp. Vis. Pattern Recog.*, 2014.

[18] V Ljosa, K L Sokolnicki, and A E Carpenter. Annotated high-throughput microscopy image sets for validation. *Nature methods*, 9(7):637–637, 2012.

[19] J Masci, U Meier, D Cireşan, and J Schmidhuber. Stacked convolutional auto-encoders for hierarchical feature extraction. In *Int. Conf. Art. Neur. Net.*, pages 52–9, 2011.

[20] S Melacci and M Belkin. Laplacian support vector machines trained in the primal. *J. Mach. Learn. Res.*, 12(Mar):1149–84, 2011.

[21] M Nardon and P Pianca. Simulation techniques for generalized gaussian densities. *Journal of Statistical Computation and Simulation*, 79(11):1317–1329, 2009.

[22] M Novey, T Adali, and A Roy. A complex generalized gaussian distribution— characterization, generation, and estimation. *IEEE Transactions on Signal Processing*, 58(3):1427–1433, 2010.

[23] L Ruff, R Vandermeulen, N Goernitz, L Deecke, S A Siddiqui, A Binder, E Müller, and M Kloft. Deep one-class classification. In *Int. Conf. Mach. Learn.*, pages 4393–402, 2018.

[24] L Ruff, RA Vandermeulen, N Görnitz, A Binder, E Müller, K-R Müller, and M Kloft. Deep semi-supervised anomaly detection. In *Int. Conf. Learn. Rep.*, 2020.

[25] T Schlegl, P Seeböck, S M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *Int. Conf. Info. Proc. Med. Imag.*, pages 146–57, 2017.

[26] B Scholkopf, J Platt, J Shawe-Taylor, A Smola, and R Williamson. Estimating the support of a high-dimensional distribution. *Neural Comp.*, 13(7):1443–71, 2001.

[27] B Scholkopf, A Smola, and K Muller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comp.*, 10(5):1299–1319, 1998.

[28] B Schölkopf, R C Williamson, A J Smola, J Shawe-Taylor, and J C Platt. Support vector method for novelty detection. In *Adv. Neural Info. Proc. Sys.*, pages 582–8, 2000.

[29] MP Shah, SN Merchant, and SP Awate. Abnormality detection using deep neural networks with robust quasi-norm autoencoding and semi-supervised learning. In *IEEE Int. Symp. Biom. Imag.*, pages 568–572, 2018.

[30] D Tax and R Duin. Support vector data description. *Mach. Learn.*, 54(1):45–66, 2004.

[31] L V Maaten and Geoffrey H. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(11), 2008.

[32] K Wadhwani and SP Awate. Controllable image generation with semi-supervised deep learning and deformable-mean-template based geometry-appearance disentanglement. *Pattern Recognition*, 118:108001, 2021.

[33] Y Xia, X Cao, F Wen, G Hua, and J Sun. Learning discriminative reconstructions for unsupervised outlier removal. In *Int. Conf. Comp. Vis.*, pages 1511–9, 2015.

[34] C Zhou and RC Paffenroth. Anomaly detection with robust deep autoencoders. In *Int. Conf. Know. Dis. Data Mining*, pages 665–74, 2017.