Neurocomputing 620 (2025) 129172

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

A semi-supervised multiscale generalized-VAE framework for one-class classification

Renuka Sharma [a,b,c,*], Suyash P. Awate [b]

[a] Commonwealth Scientific and Industrial Research Organisation (CSIRO), Data61, Brisbane, Australia
[b] Computer Science and Engineering (CSE) Department, Indian Institute of Technology (IIT) Bombay, Mumbai, India
[c] IITB-Monash Research Academy, Mumbai, India

ARTICLE INFO

Communicated by J. Zhao

Keywords:
One-class classification
Variational autoencoder
Semi-supervision
Multiscale latent space
Generalized-Gaussian
Robustness
Uncertainty

ABSTRACT

Deep-learning based approaches for *unsupervised anomaly detection* typically learn either a generative model of the inlier class or a decision boundary to encapsulate the inlier class. In addition to the training data from the inlier class, the availability of a small amount of training data from the outlier class can aid in refining the classifier model using principles of semi-supervised learning. This paper proposes a novel end-to-end deep *semi-supervised variational* framework for *one-class classification* of images, leveraging data-adaptive *generalized-Gaussian* (GG) models leading to effective modeling of distributions in both latent space and image space. The framework proposes a novel variational encoder that models a distribution on a *multiscale* (here, "scale" refers to spatial resolution) latent-space encoding, together with a generalized reparameterization scheme for the GG model's sampling at each such scale. While the multiscale latent-space helps effective feature learning at coarse and fine spatial scales, the semi-supervision helps tune the feature learning to improve separability between the inliers and the outliers. Results on several publicly available industrial-imaging and medical-imaging datasets show the benefits of our framework's novel components over existing approaches.

1. Introduction

Anomaly detection is a well-known *one-class classification* (OCC) problem, aimed at discerning whether an image exhibits anomalies or not. This process plays a pivotal role in various advanced applications, including error and corrosion detection within manufacturing industries [1–3], disease identification in the medical field [4], road anomaly detection for autonomous driving [5,6], hyperspectral imagery anomaly detection [7], and video anomaly detection [8–10] within the security domain. In OCC problem, completely characterizing the anomalies (also called outliers or abnormalities) using a finite-sized training sample is nearly infeasible because of the sample's inability to capture the large variability in the abnormal class. Hence, OCC learning relies on modeling the inlier distribution, i.e., the normal-class distribution, or its associated envelope.

Early methods for OCC use hand-crafted features within variants of density-based clustering which cannot be easily generalized to large datasets exhibiting large intra/inter-class variability and to the different setups underlying anomaly-detection tasks. The methods based on *deep neural networks* (DNNs) remove the compulsion of hand-crafting features; now, the task is to learn meaningful features from the given data to discriminate inliers from the outliers where most of the DNN-based methods [11] jointly learn a feature extractor and a classifier. In that vein, this paper focuses on DNN-based frameworks for OCC in images.

For the goal of discriminating between inliers and outliers focuses on DNN-based autoencoders trained to reconstruct inlier-class images. Methods like DRAE [11] achieve anomaly detection by analyzing the reconstruction errors of such an autoencoder; the reconstruction errors will typically be smaller on the set of inliers as compared to those on the set of outliers. This gives autoencoder models, and their intermediate features, the capability to discriminate an input into inlier and outlier classes [11]. The literature indicates that variational autoencoders (VAEs) [12] offer improvements over standard autoencoders by modeling a distribution of feature embeddings in the latent space, which effectively leads to a distribution of reconstructions as their output. Thus, similar to traditional autoencoders, the reconstruction errors and the intermediate features associated with VAE outputs can also be used to discriminate between inliers and outliers.

Another approach to feature learning is to learn the distribution of latent-space encodings of the inliers and the associated high-

* Corresponding author at: Commonwealth Scientific and Industrial Research Organisation (CSIRO), Data61, Brisbane, Australia.
E-mail address: renuka.10078@gmail.com (R. Sharma).

https://doi.org/10.1016/j.neucom.2024.129172
Received 1 April 2024; Received in revised form 4 November 2024; Accepted 7 December 2024
Available online 14 December 2024

dimensional manifold. Then, strategies for OCC can rely on the closeness of the latent-space encoding of a test datum to the learned latent-space manifold, or on the quality of the autoencoder-based reconstruction for a test datum. We propose a DNN framework that leverages *both* autoencoder-based learning (to yield good reconstructions of inlier data) and manifold learning in latent-space (to yield compact distributions of inlier data).

Traditional methods for anomaly detection rely on unsupervised learning techniques applied to unlabeled data, where the anomalous samples are not known a priori, and it is presumed that the majority (or all) of the training set comprises normal data [13]. Such unsupervised approaches face the challenge of learning to discern a large variety of anomalies unseen in the training set, and especially so when the training set is of limited size. In contrast, recent methods have started to leverage the availability of a small amount of labeled abnormal data, along with the unlabeled (assumed mostly normal) data, during training. This has led recent methods towards higher accuracy and better generalization to test data [13,14]. Indeed, in addition to the assumption of immense (and unseen) variability in the outlier class, the imbalance between the sample sizes of the inlier class and outlier class data is so large that a fully-supervised approach remains either inapplicable or grossly sub-optimal. Thus, we propose to augment our OCC framework to leverage *semi-supervised* learning that utilizes the available expert-labeled outlier data.

For image analysis, standard popular autoencoder-based approaches employ architectures that have a multiscale latent space with skip connections— in this paper, the term "scale" refers to the spatial resolution associated with an image—for example, the U-Net [15] transfers a rich set of features at multiple scales of the input image, computed by the encoder, to the decoder. Variational autoencoders (VAEs) leverage Bayesian statistical modeling to improve model learning and inference over (non-variational) autoencoders. However, standard VAEs employ architectures that either fail to supply the decoder with higher-resolution features (e.g., without skip connections) or model the latent-space distribution only at the coarsest scale. In contrast, we propose a novel *multiscale* VAE framework that (i) models distributions at multiple scales in latent-space and (ii) transfers multi-resolution information from the encoder to the decoder.

Standard VAE schemes model a Gaussian distribution in latent space, where the mean and scale (or standard deviation) parameters are output by the encoder. The motivation behind introducing GG modeling in latent space has been discussed in detail in our earlier work on the ss-gVAE [16]. The GG modeling improves the data-adaptive distributional modeling of the real-world dataset in latent space, subsuming robust modeling through the GG's shape parameters and uncertainty-aware modeling through the GG's scale parameters. In this way, GG modeling helps tackle the challenges associated with modeling real-world datasets which might not follow Gaussian models in latent space. This paper proposes to generalize the earlier strategy [16] by designing a *generalized Gaussian* (GG) model at every scale in the multiscale latent space, where the encoder additionally outputs the distributions' shape parameters. In addition, we propose a decoder model that also outputs a GG distribution on the reconstructed images. Thus, our framework proposes a data-adaptive modeling strategy that enables robust and uncertainty-aware statistical modeling in latent space as well as image space.

This paper makes novel contributions. First, we propose a novel statistical-learning framework that (i) extends a VAE to incorporate a *multiscale* latent space and (ii) leverages data-adaptive *generalized-Gaussian* (GG) models across the multiscale latent space and the image space to enable uncertainty-aware and robust statistical modeling. We call this framework as the multiscale generalized VAE (ms-gVAE). Second, to enable the learning of the ms-gVAE model using back-propagation, we propose a *reparameterization* scheme associated with the (variational) sampling in the multiscale latent space from the GG distributions. Third, we extend ms-gVAE to a semi-supervised ms-gVAE (ss-ms-gVAE) framework that can leverage some labeled outlier data during training to improve performance. Fourth, we propose to augment the novel ms-gVAE framework, for OCC with image data, with a DNN-based classifier module in latent space to improve the model's ability to separate the inlier-class manifold from the outlier class; we call this framework as c-ss-ms-gVAE. Fifth, we evaluate the proposed c-ss-ms-gVAE framework, along with many of its ablated versions and six baseline methods, on several publicly available real-world industrial-imaging and medical-imaging datasets to provide empirical insights into our methods.

The rest of the paper is organized as follows. Section 2 compares and contrasts our framework with existing methods. Section 3 details our OCC formulations (ms-gVAE, ss-ms-gVAE, and c-ss-ms-gVAE) for learning and inference. Section 4 elaborates the experiments on the industrial and medical datasets, including comparisons with several baselines and ablated versions to show the significance of novelties in our framework. Section 5 concludes the paper.

2. Related work

Early methods using OCC for anomaly detection rely on variants of kernel principal component analysis (KPCA) [17] or the support vector machine (SVM). Examples involving the latter include the one-class SVM (OC-SVM) [18] and the support-vector data description (SVDD) [19]. Such methods use hand-crafted image features and reproducing kernels. Also, the model learning in these OCC methods is unsupervised, in the sense that the learning relies solely on unlabeled training data where the majority (or all) of the training set is assumed to consist of "normal" data [13].

Another class of methods rely on hierarchical density-based clustering, incorporating unsupervised or semi-supervised learning, e.g., HDBSCAN* [20] deals with both global and local outlier detection using the GLOSH outlier measure, and also facilitates different types of visualizations that can then be turned into significant clusters. In contrast, the use of DNNs for feature extraction significantly alleviates the need for hand-crafting features, but rather allows the feature extraction to be learned along with the classifier by solving a unified optimization problem.

SVDD is a methodology for anomaly detection where the goal is to construct a compact hypersphere enveloping the feature representations of the unlabeled (inlier-class) training data in the unsupervised setup. DeepSVDD [13] extends this notion of SVDD using a two-stage process, where it first learns a DNN-based autoencoder for the inlier class, and then uses the pre-trained encoder to get the feature representations for fitting the hypersphere in the latent space. The assumption is that, during inference, the anomalous input data will lie outside this hypersphere. DASVDD [21] addresses the "hypersphere collapse" differently than DeepSVDD. Hypersphere collapse is where the network converges to the trivial solution of all-zero weights. In DASVDD the hypersphere center is a free optimization parameter unlike DeepSVDD where the hypersphere center is fixed and network biases are set to zero. Interestingly, if a sufficient number of labeled anomalies are available for training, hypersphere collapse does not pose a problem due to the conflicting objectives for labeled and unlabeled data. Some other methods for anomaly detection rely on generative adversarial networks (GANs), e.g., AnoGAN [22] that learns the latent-space distribution/manifold of the inlier-class data. During inference, AnoGAN compares the input's encoding with this learned distribution to make a prediction. Unlike our proposed framework, methods like DeepSVDD and AnoGAN work without DNN decoders, without variational learning, without a focus on robust or uncertainty-aware statistical modeling, and without semi-supervision.

Some works [11,23–25] on DNN-based OCC learn an autoencoder for the normal-class, where the autoencoder aims to reconstruct normal-class data more accurately than the abnormal-class data. Thus,

the classification relies on the magnitude of the residual in the reconstructed images being typically larger for abnormal-class images. Autoencoder-based methods like DRAE [11] typically outperform kernel-based methods like [26]. RCAE [24] proposes an inductive learning scheme that extends robust PCA using a nonlinear autoencoder. In order to incorporate robustness during model learning, the input is often corrupted by noise and/or pollution/misclassification. This intends to make the training task more challenging and the resulting model more robust to such outliers; such approaches, appearing in DRAE [11] and RCAE [24], however, do *not* reformulate the optimization problem. In contrast, our proposed framework designs a variational-learning formulation that incorporates both (i) robust statistical modeling as well as (ii) uncertainty-aware modeling, by relying on generalized-Gaussian models in latent space as well as image space.

Several generic semi-supervised methods [3,27–29] for anomaly detection are used in manufacturing industries because label generation is time-consuming and labor-intensive. This is also the case in many other real-world scenarios, including, but not limited to, the medical [30] domain, and this motivates research in the methods on semi-supervised anomaly detection. Our application scenario of OCC, and thereby our notion of semi-supervision as well, differs significantly from those in other generic semi-supervised methods like pseudo-labeling [31] and co-training [32]. In the pseudo-labeling-approach [31] based semi-supervision, the network is trained using (sufficiently large amounts of) labeled data and unlabeled data simultaneously. During training, for the unlabeled data, the pseudo-labeling approach iteratively (re)assigns pseudo-labels to the class that has the maximum predicted probability given by the current version of the model being trained. These pseudo-labels are in turn used as if they were true labels in subsequent training iterations. The co-training [32] approach implements two individual classifiers based on two views of the labeled data for assigning pseudo-labels to the unlabeled data. The views are independent of the given class. Both pseudo-labeling and co-training approaches do not focus on an OCC application scenario; in contrast, given the pseudo labels, they assume the problem to be a typical two-class or multi-class classification problem ignoring any class imbalance in the training set. Unlike the pseudo-labeling and co-training approaches, our context is of OCC where (i) we are not trying to label the unlabeled training data, but (ii) make the assumption that the majority of the training dataset consists of "normal" data [13]. Therefore, within our framework of semi-supervision, we encounter a real-world scenario where the majority of the provided data is normal, with the possibility of some contamination. Our objective is to construct a robust model capable of detecting anomalies despite these genuine challenges in the dataset with the help of some expert-labeled samples, if available. Furthermore, unlike the training and evaluation strategies in the papers on pseudo-labeling and co-training methods, our paper includes two challenging real-world evaluation scenarios where (i) the labeled training data exhibits some labeling errors (as is also done in the DeepSAD work [14]) and (ii) the training data fails to incorporate some kinds of data belonging to sub-classes of the outlier class but such data do occur during deployment (this scenario being the key motivation for the OCC approach itself).

Traditional unsupervised OCC-based anomaly detection methods only use the inlier data for training. However, with the advent of improved datasets with some available labeled data, a *small training set of expert-labeled 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 OCC methods [33–35] are able to leverage such limited information about the abnormal class using transductive learning to improve performance, typical DNN-based OCC methods [13,22–24] are unable to leverage such information. Among OCC methods, SSAD [35], DeepSAD [14], and QSSAE [36] leverage semi-supervision to learn a DNN-based one-class classifier. The semi-supervised OCC methods of SSAD [35] and DeepSAD [14] relate to SVDD that focuses on mapping normal data to a compact subspace within latent space. SSAD is a semi-supervised extension of SVDD, DeepSAD is an extension of DeepSVDD. 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. QSSAE [36] extends DCAE to leverage such limited supervision; analogous to RCAE, QSSAE aims to be robust to the corruption in the training data. The evaluation strategy in the DeepSAD work [14] underscores the benefits of semi-supervision in the case of learning models from training data that contains some misclassification/pollution in the unlabeled/normal class; we also follow that strategy. Generic uncertainty-aware modeling approaches [37–39] are designed to handle real-world datasets susceptible to inconsistencies. The goal is to develop robust models capable of addressing these inevitable challenges, which are particularly significant in high-impact domains such as medicine and manufacturing. Several methods [38,39] have been effectively applied to disease prediction, yielding highly confident results. Deep multiscale CNN (DMSCNN) [28] provide robust statistical modeling of the feature space for tool-wear prediction. In our case, when the training set involves pollution, gVAE-based learning is unsupervised in the sense that it does not know inliers from outliers. In contrast, semi-supervised learning within ss-gVAE (and its derivatives) leverages some data explicitly labeled as outliers. Unlike DeepSAD, our framework proposes (i) a unified optimization framework that relies on (ii) learning image reconstructions for the inlier class using an autoencoder as well as (iii) learning a decision boundary in latent-space using a DNN-based classifier. Moreover, unlike DeepSAD and QSSAE, we propose a (i) variational framework that (ii) incorporates robust statistical and uncertainty-aware modeling.

This paper significantly extends our preliminary work [16,30], theoretically and empirically. First, this paper extends the theoretical frameworks in [16,30] by proposing a *multiscale latent space* within a VAE statistical framework. This new framework brings together the best of (i) the U-Net architecture's skip connections across image scales/resolutions [15] and (ii) variational learning in the VAE [12], extending them both within a sound statistical framework involving robust statistical modeling and uncertainty-aware learning using generalized-Gaussian models in the latent space and the image space. Indeed, it improves over the results in our preliminary work during both unsupervised learning and semi-supervised-learning. Second, unlike our preliminary work, this paper introduces a *multiscale classifier* module in latent space to improve the learning of the manifold of inlier-class multiscale encodings and help separate the outlier-class encodings. Third, this paper provides detailed theoretical descriptions, and comprehensive empirical analysis (quantitative and qualitative) on a *larger number of datasets* (publicly available) including three medical-imaging datasets and ten industrial-imaging datasets. Fourth, this paper provides more empirical insights into the working of our framework using *more ablated versions*, compared to our preliminary work. Fifth, this paper also provides more insights into the performance of our methods and baselines through qualitative analysis incorporating *visualization* of (i) latent-space manifolds and (ii) image patches that exhibit various levels of challenges in classification (see Table 1).

3. Methods

For OCC of images or image regions/patches, we propose a novel multiscale generalized-VAE (ms-gVAE) statistical model and architecture (Section 3.1) and the associated theoretical formulations for learning (Section 3.2) and inference. Here, the term "scale" refers to the spatial resolution of the image. Each pooling layer in ms-gVAE helps reduce spatial resolution to allow the subsequent convolution kernels to accumulate information in successively larger local spatial neighborhoods. ms-gVAE's encoder and decoder (Fig. 1) output input/datum-specific GG distributions in multiscale latent space and

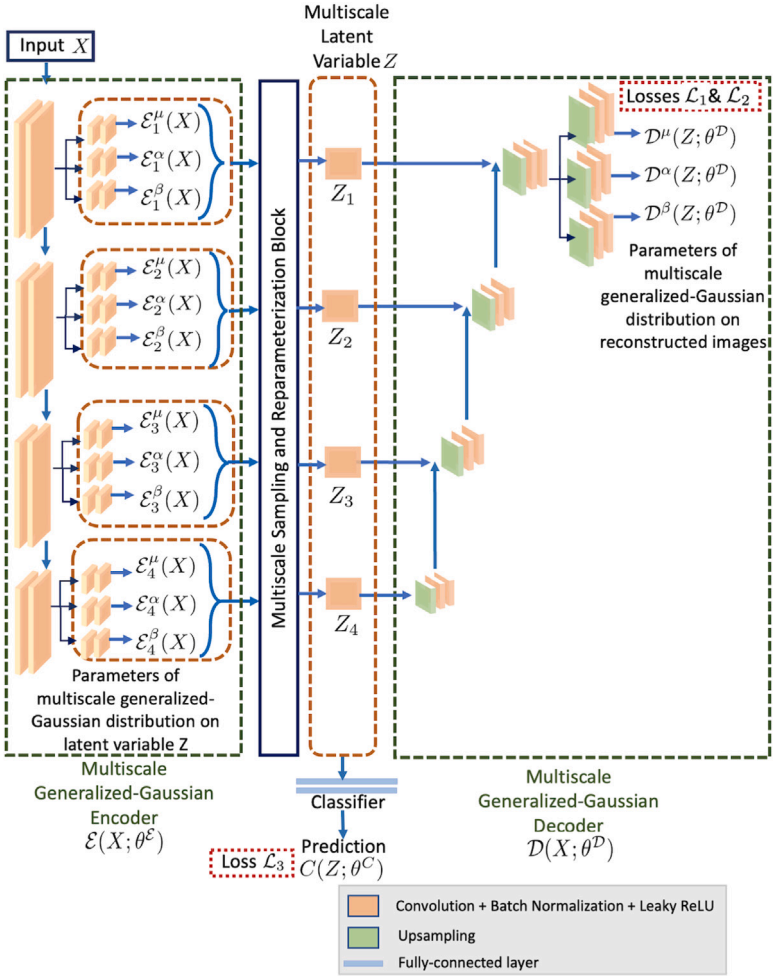


Fig. 1. A semi-supervised multiscale generalized VAE framework for OCC. Input image X processed through *multiscale encoder* $\mathcal{E}(\cdot;\theta^{\mathcal{E}})$ outputs a distribution over the *multiscale latent/hidden variable* $Z := [Z_1, Z_2, \ldots, Z_S]$. The *decoder* $\mathcal{D}(\cdot;\theta^{\mathcal{D}})$ maps the multiscale features modeled by Z to a generalized-Gaussian distribution on the reconstructed images. The classifier $C(\cdot;\theta^{C})$ maps the multiscale latent variable Z to a probability $\in (0,1)$ of belonging to the normal-class. Please refer to the symbol table Table 1 for notations.

image space. Section 3.3 proposes a latent-space GG reparameterization to enable backpropagation at any scale. Section 3.4 proposes a semi-supervised ms-gVAE (ss-ms-gVAE) framework (Fig. 1) that leverages a small training set of labeled outliers.

3.1. A multiscale generalized VAE (ms-gVAE) statistical model

Let X be a random field modeling a distribution on input images. Let Z be the *multiscale hidden/latent* random vector that captures model's compact information at multiple scales, needed to reconstruct image X. Let a DNN-based *encoder* mapping at spatial scale/resolution s be $\mathcal{E}_s(\cdot)$, which maps the input X to $\mathcal{E}_s(X)$ that models a distribution on the latent-space encoding at scale s parameterized by the (i) mean $\mathcal{E}_s^{\mu}(X)$ parameter, (ii) the scale $\mathcal{E}_s^{\alpha}(X)$ parameter, and (iii) the shape $\mathcal{E}_s^{\beta}(X)$ parameter. Thus, across S spatial scales/resolutions, the multiscale distribution on the multiscale latent-space encoding is given by $\mathcal{E}(X) := [\mathcal{E}_1(X), \mathcal{E}_2(X), \ldots, \mathcal{E}_S(X)]$, where $\mathcal{E}_s(X) := [\mathcal{E}_s^{\mu}(X), \mathcal{E}_s^{\alpha}(X), \mathcal{E}_s^{\beta}(X)]$. In our variational framework, at scale s, an instance of the latent-space encoding Z_s can be obtained by sampling from the GG distribution parameterized by $\mathcal{E}_s(X)$. Thus, an instance of the multiscale encoding Z can be sampled from the multiscale GG distribution parameterized by $\mathcal{E}(X)$. Let a DNN-based multiscale *decoder* model a mapping $\mathcal{D}(\cdot;\theta^{D})$, parameterized by θ^{D}, which maps the multiscale latent-space encoding Z to the parameters $\mathcal{D}(Z;\theta^{D})$ of a GG distribution on the autoencoder-reconstructed images. Thus, within our framework, the encoder and decoder both model GG distributions (respectively, over the multiscale latent-space and the image space).

3.2. Unsupervised learning with ms-gVAE

In unsupervised learning, all the unlabeled training data is considered normal and belongs to the set of normal data $\{X_n\}_{n=1}^{N}$. Let $\mathcal{G}(\cdot;\mu,\alpha,\beta)$ be the factorized GG with mean vector μ, a vector of log-scale parameters α, and a vector of log-shape parameters β. Similar, in spirit, to the VAE learning strategy [12], ms-gVAE learning optimizes $\theta^{\mathcal{E}\mathcal{D}}$ by minimizing the loss $\mathcal{L}_1$ [16] on the unlabeled training-set images $\{X_n\}_{n=1}^{N}$. For each X_n, let z^{ni} be the ith independent draw of the multiscale latent-space vector from the PDF $Q(Z|X_n;\theta^{\mathcal{E}})$. Let the notation $z^{ni}(X_n,\theta^{\mathcal{E}})$ indicate that z^{ni} depends on encoder parameters $\theta^{\mathcal{E}}$ and X_n. We propose to estimate the ms-gVAE parameters $\theta^{\mathcal{E}\mathcal{D}} := \theta^{\mathcal{E}} \cup \theta^{D}$ by maximizing the likelihood of the observed training-set images which is equivalent to minimizing the loss

$$\mathcal{L}_1(\{X_n\}_{n=1}^{N};\theta^{\mathcal{E}\mathcal{D}}) := \frac{1}{NI}\sum_{n=1}^{N}\sum_{i=1}^{I}[0.5\|z^{ni}(X_n,\theta^{\mathcal{E}})\|_2^2 + \log\mathcal{G}(z^{ni}(X_n,\theta^{\mathcal{E}});\mathcal{E}^{\mu}(X_n;\theta^{\mathcal{E}}),\mathcal{E}^{\alpha}(X_n;\theta^{\mathcal{E}}),\mathcal{E}^{\beta}(X_n;\theta^{\mathcal{E}})) - \log\mathcal{G}(X_n;\mathcal{D}^{\mu}(z^{ni}(X_n,\theta^{\mathcal{E}});\theta^{\mathcal{D}}),\mathcal{D}^{\alpha}(z^{ni}(X_n,\theta^{\mathcal{E}});\theta^{\mathcal{D}}),\mathcal{D}^{\beta}(z^{ni}(X_n,\theta^{\mathcal{E}});\theta^{\mathcal{D}}))]. \quad (1)$$

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 dth scalar component of its vector argument. Then, $[\log\mathcal{G}(b;\mu,\alpha,\beta)]_d := \log((\tau+\exp([\beta]_d))/(\Delta+\exp([\alpha]_d))) - (((b-[\mu]_d)/(\Delta+\exp([\alpha]_d)))^2+\epsilon)^{0.5(\tau+\exp([\beta]_d))} - \log(\Gamma(1/(\tau+\exp([\beta]_d)))) +$ constant, where Δ, ϵ, and τ (all $\in \mathbb{R}^+$) act as regularizers.

Table 1
Symbol table: Explanations for the notations used for c-ss-ms-gVAE.

Symbol	Explanation
$\mathcal{X}$	Input space; random fields are extracted from this space
X	Input image/random field for normal samples, every image is represented as a random field
Y	Input image/random field for abnormal samples
Z_i	Latent vector at scale $\mathbf{i}$, sampled from the encoded outputs (mean $\mathcal{E}_i^{\mu}$, scale $\mathcal{E}_i^{\alpha}$, and shape $\mathcal{E}_i^{\beta}$) in the latent space
Z	For S scales, the multiscale latent/hidden variable $Z := [Z_1, Z_2, \ldots, Z_S]$.
$\mathcal{E}(\cdot;\theta^{\mathcal{E}})$	Encoder for gVAE [16] which maps the input (X or Y) to the latent variable (Z) through the components: mean $\mathcal{E}^{\mu}(.;\theta^{\mathcal{E}})$, scale $\mathcal{E}^{\alpha}(.;\theta^{\mathcal{E}})$, and shape $\mathcal{E}^{\beta}(.;\theta^{\mathcal{E}})$
$\mathcal{D}(\cdot;\theta^{\mathcal{D}})$	Decoder for gVAE [16] which maps the latent variable (Z) to the reconstructions (mean $\mathcal{D}^{\mu}(Z;\theta^{\mathcal{D}})$, scale $\mathcal{D}^{\alpha}(Z;\theta^{\mathcal{D}})$, and shape $\mathcal{E}^{\beta}(X;\theta^{\mathcal{E}})$)
$C(\cdot;\theta^{C})$	Classifier for c-ss-ms-gVAE, operating at the multiscale latent-variable, Z
$\theta^{\mathcal{E}}$	Parameters of the encoder $\mathcal{E}(\cdot;\theta^{\mathcal{E}})$
$\theta^{\mathcal{E}_i^{\mu}}$, $\theta^{\mathcal{E}_i^{\alpha}}$, and $\theta^{\mathcal{E}_i^{\beta}}$	GG-mean, GG-scale, and GG-shape parameters of the encoder $\mathcal{E}(\cdot;\theta^{\mathcal{E}})$ at scale $\mathbf{i}$
$\theta^{\mathcal{D}}$	Parameters of the decoder $\mathcal{D}(\cdot;\theta^{\mathcal{D}})$
$\theta^{\mathcal{D}^{\mu}}$, $\theta^{\mathcal{D}^{\alpha}}$, and $\theta^{\mathcal{D}^{\beta}}$	GG-mean, GG-scale, and GG-shape parameters of the decoder $\mathcal{D}(\cdot;\theta^{\mathcal{D}})$
θ^{C}	Parameters of the classifier $\mathcal{D}(\cdot;\theta^{C})$
$\theta^{\mathcal{ED}} := \theta^{\mathcal{E}} \cup \theta^{\mathcal{D}}$	c-ss-ms-gVAE parameters; combination of the parameters for encoders and decoders
$P(X, Z;\theta^{\mathcal{ED}})$	PDF of the complete data, models the dataset (parameterized by the autoencoder parameters)
$P(Z\|X;\theta^{\mathcal{ED}})$	True latent-variable posterior PDF
$\log P(X;\theta^{\mathcal{ED}})$	log likelihood of the input X given the autoencoder parameters
$Q(Z\|X;\theta^{\mathcal{E}})$	Conditional PDF of the latent-space encoding Z
$G(.;\mu,c,\rho)$	The univariate generalized-Gaussian with mean μ, scale c, and shape ρ

3.3. Reparameterizing the generalized Gaussian

To enable backpropagation-based optimization for encoder parameters $\theta^{\mathcal{E}}$, we propose to reparameterize the ith independent draw at any scale in the multiscale autoencoder-based scheme $z^{ni} \sim \mathcal{G}(Z;\mathcal{E}^{\mu}(X_n;\theta^{\mathcal{E}}), \mathcal{E}^{\alpha}(X_n;\theta^{\mathcal{E}}), \mathcal{E}^{\beta}(X_n;\theta^{\mathcal{E}}))$ using a hierarchical (2-level) reparameterization scheme. Let Gamma(a,b) be the Gamma PDF with shape parameter a and scale parameter b. First, we reparameterize [40] z^{ni} based on Gamma random variables as

$$[z^{ni}]_d := [\mathcal{E}^{\mu}(X_n;\theta^{\mathcal{E}})]_d + (\Delta + [\exp(\mathcal{E}^{\alpha}(X_n;\theta^{\mathcal{E}}))]_d) B^{nid} |Y^{nid}|^{1/(\tau+\exp([\mathcal{E}^{\beta}(X_n;\theta^{\mathcal{E}})]_d))}, \quad (2)$$

where random variables Y^{nid} have the PDF Gamma$(1/(\tau+[\mathcal{E}^{\beta}(X;\theta^{\mathcal{E}})]_d), 1)$ and B^{nid} take one of the values in $\{-1,+1\}$ with probability 0.5. Second, we leverage implicit reparameterization gradients [41] for the Gamma random variables Y^{nid}.

3.4. Semi-supervised ms-gVAE learning with ss-ms-gVAE

To improve OCC learning over ms-gVAE, we extend gVAE to propose ss-ms-gVAE that is able to leverage a small set of expert-labeled inliers in the existing normal set $\{X_l\}_{l=1}^{L}$ (from unsupervised learning) along with some expert-labeled outliers $\{Y_m\}_{m=1}^{M}$, where typically $M \ll N + L$. While ms-gVAE learning seeks higher values of the log-likelihoods $\log P(X_n;\theta^{\mathcal{ED}})$ by seeking higher values of the associated functionals $\mathcal{F}(Q(Z|X_n;\theta^{\mathcal{E}});\theta^{\mathcal{ED}})$, ss-ms-gVAE learning additionally seeks low values of the log-likelihoods $\log P(Y_m;\theta^{\mathcal{ED}})$ of the labeled outliers Y_m. Analogous to $\mathcal{L}_1(\cdot)$ for ms-gVAE, we propose another loss term leveraging the expert-labeled normal and abnormal-class training subsets $\{X_l\}_{l=1}^{L}$ and $\{Y_m\}_{m=1}^{M}$ as

$$\begin{aligned}&\mathcal{L}_2(\{X_l\}_{l=1}^{L}, \{Y_m\}_{m=1}^{M};\theta^{\mathcal{ED}}) := \\ &\mathcal{L}_1(\{X_l\}_{l=1}^{L};\theta^{\mathcal{ED}}) - \frac{1}{MJ}\sum_{m=1}^{M}\sum_{k=1}^{K}[0.5\|z^{mk}(Y_m,\theta^{\mathcal{E}})\|_2^2 + \\ &\log \mathcal{G}(z^{mk}(Y_m,\theta^{\mathcal{E}});\mathcal{E}^{\mu}(Y_m;\theta^{\mathcal{E}}),\mathcal{E}^{\alpha}(Y_m;\theta^{\mathcal{E}}),\mathcal{E}^{\beta}(Y_m;\theta^{\mathcal{E}})) - \\ &\log \mathcal{G}(Y_m;\mathcal{D}^{\mu}(z^{mk}(Y_m,\theta^{\mathcal{E}});\theta^{\mathcal{D}}),\mathcal{D}^{\alpha}(z^{mk}(Y_m,\theta^{\mathcal{E}});\theta^{\mathcal{D}}),\mathcal{D}^{\beta}(z^{mk}(Y_m,\theta^{\mathcal{E}});\theta^{\mathcal{D}}))].\end{aligned} \quad (3)$$

For each input Y_m, the latent-space vector $z^{mj}(Y_m,\theta^{\mathcal{E}})$ is the jth independent draw from $Q(Z|Y_m;\theta^{\mathcal{E}}) = \mathcal{G}(Z;\mathcal{E}^{\mu}(Y_m;\theta^{\mathcal{E}}),\mathcal{E}^{\alpha}(Y_m;\theta^{\mathcal{E}}),\mathcal{E}^{\beta}(Y_m;\theta^{\mathcal{E}}))$, and is reparameterized using the scheme in Section 3.3.

Thus, ss-ms-gVAE learning minimizes $\mathcal{L}_1(\{X_n\}_{n=1}^{N};\theta^{\mathcal{ED}}) + \lambda_2\mathcal{L}_2(\{X_l\}_{l=1}^{L},\{Y_m\}_{m=1}^{M};\theta^{\mathcal{ED}})$, where $\lambda_2 > 0$ is a free parameter.

3.5. Augmenting ss-ms-gVAE with a classifier in latent space (c-ss-ms-gVAE)

Within the semi-supervised learning mode, i.e., when we have access to the expert-labeled (small) training subset, we propose an additional loss term $\mathcal{L}_3(\cdot)$ to promote the separability of the normal-class and abnormal-class distributions in latent space. To do so, we propose to introduce a nonlinear classifier $C(\cdot;\theta^{C})$ to discriminate between the distributions of the latent-space encodings of the independent draws $\cup_{n=1}^{N}\cup_{i=1}^{I}\{z^{ni}\}$, $\cup_{l=1}^{L}\cup_{j=1}^{J}\{z^{lj}\}$, and $\cup_{m=1}^{M}\cup_{k=1}^{K}\{z^{mj}\}$, where z^{ni}, z^{lj} and z^{mk} are independent draws from multiscale latent-space PDFs at each scale, as defined earlier, and reparameterized using the scheme in Section 3.3. For a latent-space encoding z, the classifier output $C(z;\theta^{C})$ indicates the probability $\in (0,1)$ of encoding z belonging to the normal class. Thus, we propose the associated loss term $\mathcal{L}_3(\cdot)$ as the cross-entropy loss

$$\begin{aligned}&\mathcal{L}_3(\{X_n\}_{n=1}^{N},\{X_l\}_{l=1}^{L},\{Y_m\}_{m=1}^{M};\theta^{\mathcal{ED}},\theta^{C}) := \\ &-\frac{1}{MJ}\sum_{m=1}^{M}\sum_{j=1}^{J}\log(C(z^{mj}(Y_m,\theta^{\mathcal{E}});\theta^{C})) \\ &-\frac{1}{NJ}\sum_{n=1}^{N}\sum_{j=1}^{J}\log(1-C(z^{ni}(X_n,\theta^{\mathcal{E}});\theta^{C})).\end{aligned} \quad (4)$$

In addition to ss-ms-gVAE, the classifier-augmented version c-ss-ms-gVAE also promotes good separability of the distributions of latent-space encodings of the normal-class and abnormal-class training sets.

Thus, c-ss-ms-gVAE learning minimizes

$$\begin{aligned}&\mathcal{L}_1(\{X_n\}_{n=1}^{N};\theta^{\mathcal{ED}}) + \lambda_2\mathcal{L}_2(\{X_l\}_{l=1}^{L},\{Y_m\}_{m=1}^{M};\theta^{\mathcal{ED}}) \\ &+\lambda_3\mathcal{L}_3(\{X_n\}_{n=1}^{N},\{X_l\}_{l=1}^{L},\{Y_m\}_{m=1}^{M};\theta^{\mathcal{ED}},\theta^{C}),\end{aligned} \quad (5)$$

where $\lambda_2 > 0$ and $\lambda_3 > 0$ are free parameters.

3.6. Inference strategy

We propose a generic inference strategy that holds in both unsupervised and semi-supervised learning scenarios. During all forms of ms-gVAE based learning, i.e., (i) in the ms-gVAE in the unsupervised learning mode as well as (ii) in the extensions ss-ms-gVAE and c-ss-ms-gVAE that exist only in the semi-supervised learning mode, the following observations hold. First, for inliers X, the KL$(Q(Z|X;\theta^{\mathcal{E}}) \| P(Z))$ term promotes $Q(Z|X)$ to be close to $P(Z)$, thereby promoting the multiscale generalized-Gaussian encoded mean $\mathcal{E}^{\mu}(X;\theta^{\mathcal{E}})$ to be close to the origin. Also, the multiscale generalized-Gaussian decoder-based loss $-\log P(X|Z;\theta^{\mathcal{D}})$ promotes the multiscale decoder-output

Algorithm 1 Algorithm for c-ss-ms-gVAE

Input: Training set which contains unlabeled samples $\{X_n\}_{n=1}^{N}$, expert-labeled normal samples $\{X_l\}_{l=1}^{L}$, and expert-labeled anomalous samples $\{Y_m\}_{m=1}^{M}$. Expert-labeled samples are only available during semi-supervision.

Parameters: Generalized Gaussian (GG) encoder parameters at each scale (in the multi-scaled architecture) i: $\theta^{\mathcal{E}_i^{\mu}}$ (mean), $\theta^{\mathcal{E}_i^{\alpha}}$ (scale), and $\theta^{\mathcal{E}_i^{\beta}}$ (shape). GG decoder parameters are $\theta^{\mathcal{D}^{\mu}}$ (mean), $\theta^{\mathcal{D}^{\alpha}}$ (scale), and $\theta^{\mathcal{D}^{\beta}}$ (shape).

Output: AUC score for the test set (contains normal samples, anomalous samples, and some novel-anomaly test samples from the anomalous categories not introduced in the training set during semi-supervision).

Initialization (warm-start): Parts of DNN acting on scale and shape are kept frozen in order to train for each parameter (GG-mean μ, GG-scale α, and GG-shape β) in a step-wise manner.

1. Initialize only GG-mean μ parameters for both encoder and decoder by minimizing the reconstruction errors. This is done by keeping GG-scale α and GG-shape β constant, at 1 and 2 respectively to simulate Gaussian distribution.

2. Learn GG-mean μ and GG-scale parameters α while keeping GG-shape β at a constant 2 by optimizing the learning objective (Equation (5)) for μ and α.

For each training batch B do:

1. Train the entire network (encoder and decoder parameters at all scales) using the initialized weights.
2. Optimize the learning objective (Equation (5)) to reduce the loss on normal images (by reducing $\mathcal{L}_1$), increase the loss on anomalous images (by increasing $\mathcal{L}_2$), and simultaneously reducing the classification loss ($\mathcal{L}_3$).

Inference: For a test sample u, compute the anomaly score $S(u)$ (Section 3.6). If $S(u)$ is greater than the tuned threshold (Section 3.6), then label it as anomalous, else label it as normal.

mean $D^{\mu}(Z;\theta^{D})$ to be close to the inlier X. Analogously, for the outliers Y (when they are provided during learning), the objective function promotes (i) the encoded mean $\mathcal{E}^{\mu}(Y;\theta^{\mathcal{E}})$ to be away from the origin, (ii) the multiscale latent-encoding distribution to be separated from that of inliers, and (iii) the multiscale decoder-output mean $D^{\mu}(Z;\theta^{D})$ to be far from the input Y. Thus, irrespective of whether a model is learned in unsupervised mode or semi-supervised mode, for a test image U, inference can rely on the multiscale latent-space encoding (e.g., $\mathcal{E}^{\mu}(U;\theta^{\mathcal{E}})$) and/or decoder output (e.g., $D^{\mu}(Z;\theta^{D})$). For test image U, we propose an anomaly score as $S(U) := \|\mathcal{E}^{\mu}(U;\theta^{\mathcal{E}})\|_2$. We find that, typically, the outliers lead to scores that are larger as compared to the inliers. We classify U as normal or anomalous based on a threshold on the score, where the threshold is a free parameter that can be tuned using an appropriate validation set.

3.7. Optimization strategy during training

We optimize our DNN hierarchically as follows. First, we ignore the parameters θ^{C} (that would exist solely in the semi-supervised mode) and we ignore the parameters corresponding to the encoder and decoder branches outputting the log-scale and log-shape parameters, fixing all scale parameters to 1 and all shape parameters to 2. In this, non-variational setting, we consider the mean-vector output as the mode-approximation to the GG distribution, and train the DNN using only the unlabeled (assumed inlier) data. Second, with this as a warm start, we now include the parameters corresponding to the log-scale branches of the network (for the encoder and the decoder), fix all shape parameters to 2, and then re-train the network. Third, we now include the parameters corresponding to the log-shape branches of the network (for the encoder and the decoder) and then re-train the network. We now have the complete ms-gVAE trained solely using inlier data. Fourth, if a small training set of expert-labeled inliers and outliers are available, then we warm start with the trained ms-gVAE, and re-train the DNN to produce the ss-ms-gVAE model. Fifth, using the trained ss-ms-gVAE model as a warm start, we now include the parameters θ^{C} corresponding to the classifier, and re-train the DNN to produce the c-ss-ms-gVAE model. In the semi-supervised learning mode, i.e., for c-ss-ms-gVAE and ss-ms-gVAE, we tune the free parameters λ_2 and/or λ_3 using a small validation set comprising inliers and outliers. We use Adam [42] optimizer with batch size 128 and weight decay $\lambda = 5 \times 10^{-7}$.

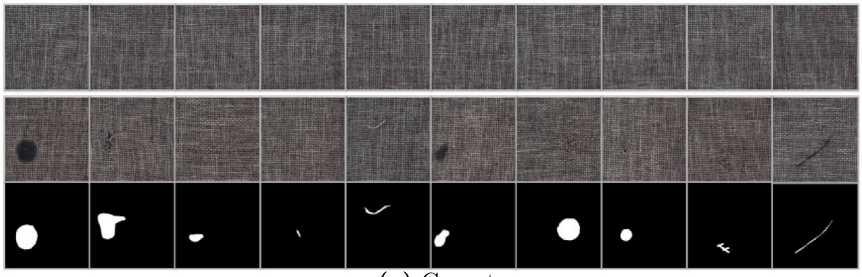
4. Results and discussion

Methods. We compare our method c-ss-ms-gVAE with seven other *baseline methods* from the literature, i.e., (i) DeepSAD [14]: semi-supervised DNN-based OCC using both unlabeled and labeled samples for training; (ii) SSAD-Hybrid [43]: semi-supervised kernel-based OCC extending SSAD [35] using pre-extracted autoencoder-based features; (iii) DeepSVDD [13]: unsupervised DNN-based OCC using only unlabeled samples for training; (iv) OC-SVM-Hybrid [43,44]: unsupervised kernel-based OCC extending OC-SVM [18] using pre-extracted autoencoder-based features; (v) ss-DCAE: semi-supervised version of DNN-based OCC [23] relying on image reconstruction using an autoencoding strategy; (vi) BinClass: fully-supervised DNN-based binary classifier using both inliers and outliers for training; and (vii) ss-gVAE, i.e., our preliminary work [16] using semi-supervised generalized-VAE using GG modeling in the latent space and image space; essentially, ss-gVAE is devoid of the multiscale latent space and the classifier present in c-ss-ms-gVAE. We evaluate each method at different levels of supervision $\gamma := M/(L+M+N)$ ranging within $[0, 0.2]$; indeed, BinClass cannot perform when $M = 0$ (i.e., $\gamma = 0$). In this paper, all kernel-based methods use the radial-basis-function kernel. For quantitative evaluation, we use the area under the receiver-operating-characteristics curve (AUC).

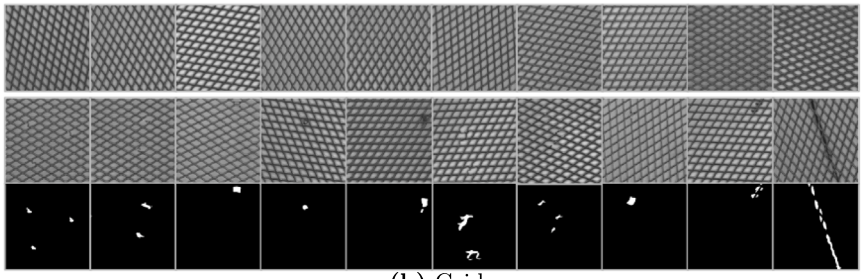
Datasets. For empirical evaluation, we use (i) ten industrial-image datasets from the MVTec repository [45] and (ii) three medical-image datasets, i.e., microscopy images involving malaria-infected blood cells [46], microscopy images involving nanofibres [47], and breast ultrasound images [48]. We split each dataset into three mutually-exclusive and exhaustive subsets for (i) training, (ii) validation (to tune free parameters), and (iii) testing. Akin to the learning and evaluation schemes motivated in the literature [11,14,24] as described in Section 2, to make the learning more challenging and induce robustness in all the learned models, we introduce 10% misclassification in the training set.

4.1. DNN architecture for c-ss-ms-gVAE

Fig. 1 illustrates the DNN architecture described in Section 3, comprising a multiscale encoder, a decoder, and a classifier. The variational learning framework motivates a sampling and reparameterization block at each scale of the multiscale latent space. Thus, each scale in the c-ss-ms-gVAE architecture consists of the encoder (feature map) block, sampling and reparameterization block, and the decoder (reconstruction block). Successive encoder stages reduce the spatial resolution, while the corresponding decoder stages incorporating spatial upsampling. The latent space incorporates features at multiple spatial resolutions, all of which are input to the decoder and the classifier. Each encoder block consists of two convolutional blocks, consisting of a convolution layer (3×3 filters; padding of 1), batch normalization, and leaky ReLU. This encoder block then forks into three sub-modules to produce the mean, scale, and shape parameters for the distribution on latent-space

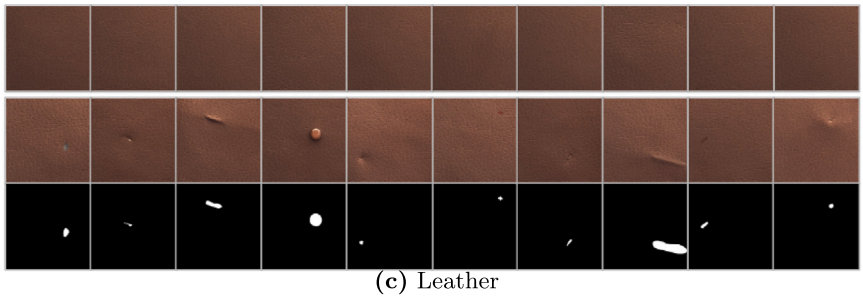


(a) Carpet

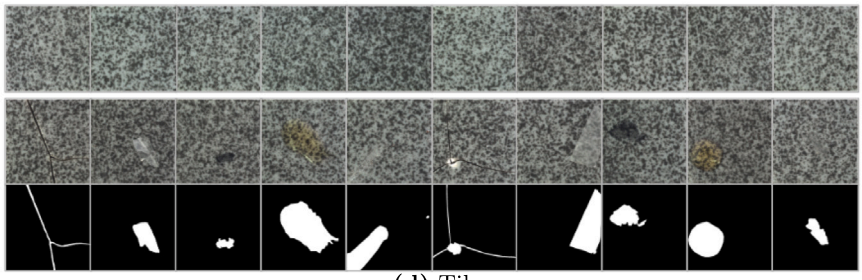


(b) Grid

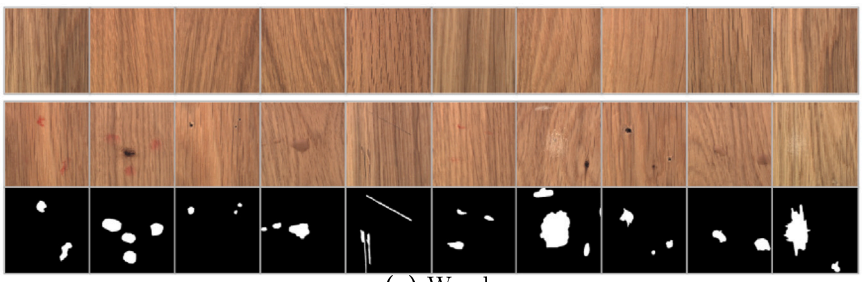
Fig. 2. MVTec dataset: Texture categories. Top Row: Normal examples. **Middle Row:** Abnormal examples. **Bottom Row:** Segmentation maps corresponding to middle row.



(c) Leather



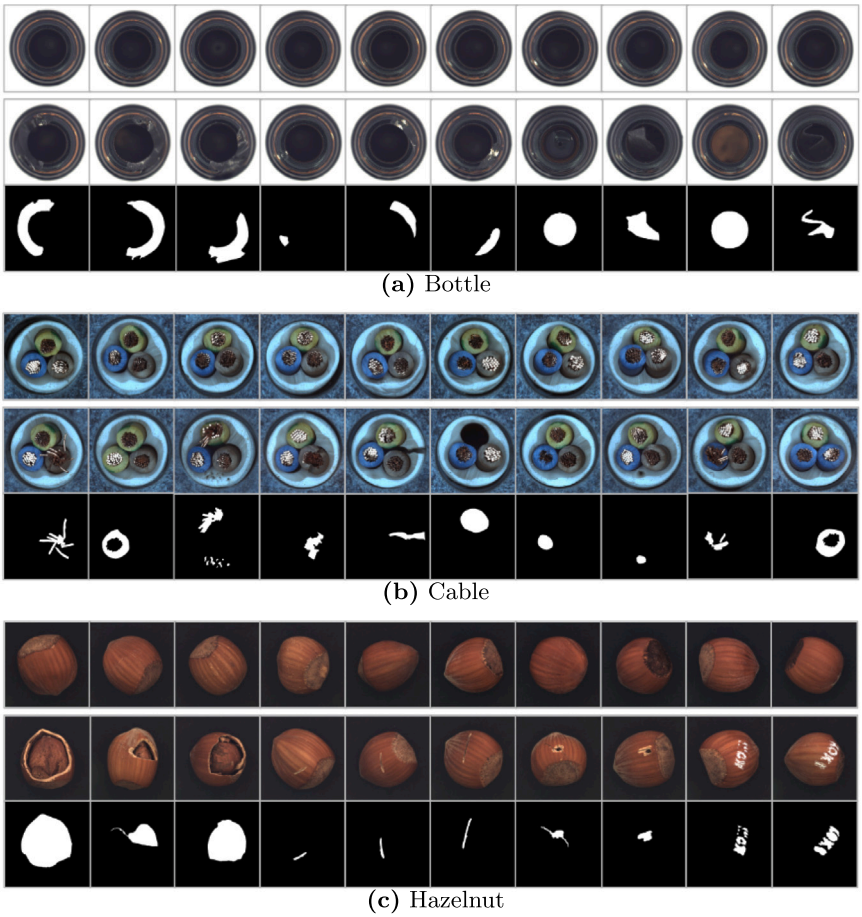
(d) Tile



(e) Wood

Fig. 2. (*continued*).

features. This forking involves three convolutional blocks consisting of a convolution layer (1×1 filters; no padding), batch normalization, and leaky ReLU. Each decoder block consists of a spatial-upsampling layer (scale factor of 2) followed by two convolutional blocks consisting of convolution layer (3×3 filters; padding of 1) followed by batch normalization and leaky-ReLU activation. These features are then passed to the next decoder block at a finer spatial scale. The final decoder block forks into three sub-modules to give the mean, scale, and shape parameters of a distribution on image reconstructions. This final fork operation involves three convolutional blocks, each comprising



(a) Bottle

(b) Cable

(c) Hazelnut

Fig. 3. MVTec dataset: Object categories. Top Row: Normal examples. **Middle Row:** Abnormal examples. **Bottom Row:** Segmentation maps corresponding to middle row.

a convolution layer (1×1 filters; no padding), batch normalization, and leaky ReLU. The multiscale latent-space encodings Z feed into the classifier block comprising two fully-connected layers.

4.2. Ablated versions of our framework

To gain insights into various components of our method, we perform the following *ablation studies* of our method c-ss-ms-gVAE detailed as follows. (i) **A0**: removes the multiscale latent space from c-ss-ms-gVAE, leaving it with the GG VAE based architecture with a classifier in the coarsest-scale latent space (similar to c-ss-gVAE [30]); (ii) **A1**: removes from A0 the latent-space components of the classifier, GG modeling, and variational learning, but includes a latent-space loss of the squared norm of the encoding from the origin (similar to the approach underlying SVDD-based methods); (iii) **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; (iv) **A3**: removes from A2 the squared-norm based loss in latent space, thereby making A3 similar to *ss-DCAE*; and (v) **A4**: removes from A2 the decoder-based loss term in image space, thereby making A4 similar to *DeepSAD*.

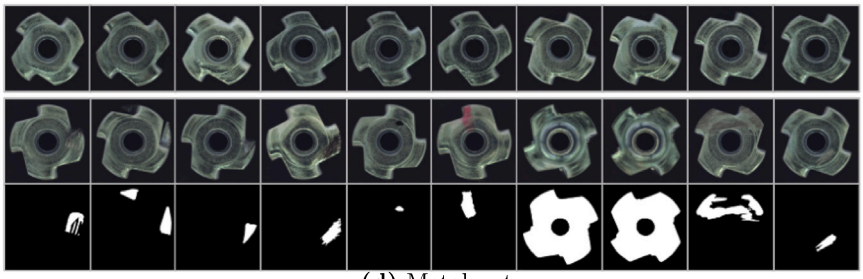
4.3. Results I: Ten industrial texture image datasets (MVTec)

The industrial inspection MVTec dataset [45] is designed for anomaly detection and includes over 5000 high-resolution images of the texture categories (Fig. 2) and object categories (Fig. 3). Each texture and object category is subdivided further into 3–8 sub-categories. 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 abnormality. We label a patch as anomalous if the corresponding anomaly mask contains at least 5% of anomalous pixels. We then resize the patches to 32×32 pixels. During semi-supervised learning, to create the expert-labeled training set comprising outliers, we use images from about half of the sub-categories. Our test set contains images from all the sub-categories to be classified into normal or abnormal; thus, the test set contains outlier classes that are unavailable during training, which is exactly the scenario motivating OCC approaches. Our DNN architecture has $S = 4$ scales.

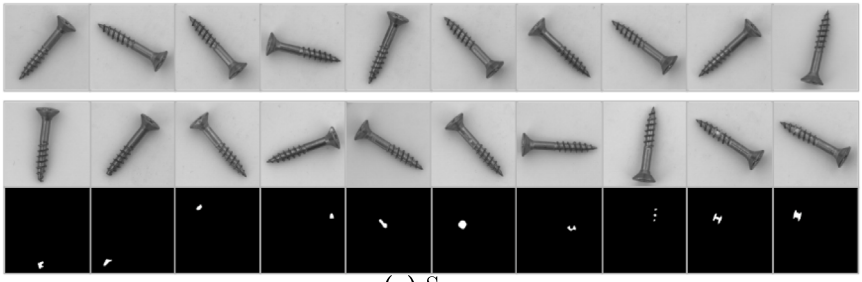
For the carpet category (Fig. 4(a)), the results show that our c-ss-ms-gVAE performs better than all other methods at virtually every level of supervision. On introducing semi-supervision during learning, SSAD-Hybrid improves upon its unsupervised version (OC-SVM-Hybrid). DeepSAD also improves over its unsupervised version (DeepSVDD) with the help of semi-supervision.

Fig. 4(b) shows that the ablated versions that rely either solely on the reconstruction-based losses (i.e., A3 that is akin to ss-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. A0, i.e., the ablated version of our proposed c-ss-ms-gVAE, replacing the multiscale latent variable by a single-scale latent variable, is unable to perform as good as c-ss-ms-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.

Table 2 shows the comparison of our methods and baselines at $\gamma = 0.2$ level of supervision. While the DNN based methods (ours, ss-gVAE, and DeepSAD) perform better than the non-DNN based ones (OC-SVM-Hybrid, SSAD-Hybrid) on the ten MVTec datasets, SSAD-Hybrid performs competitively with respect to ss-DCAE that uses the

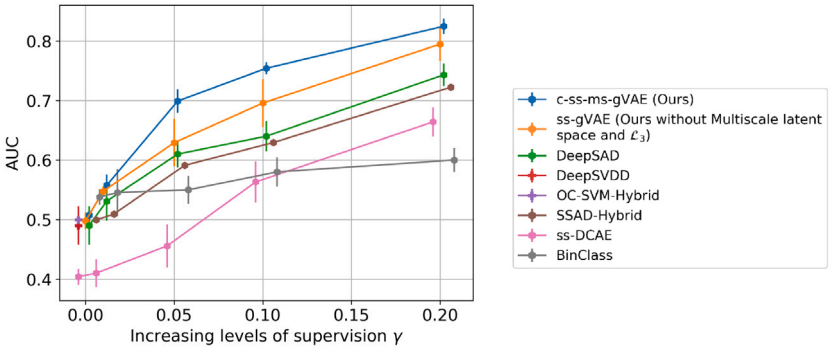


(d) Metal nut

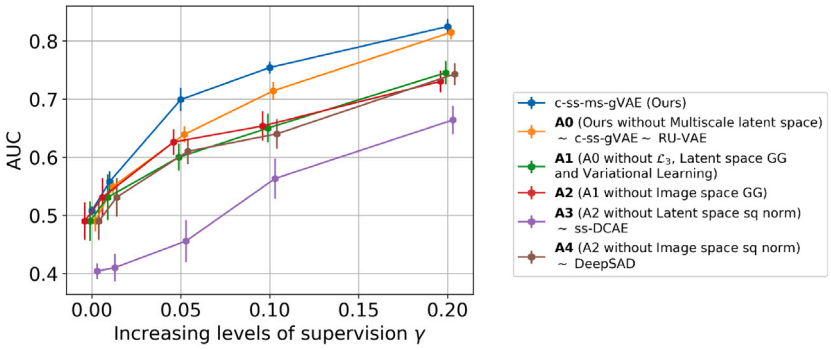


(e) Screw

Fig. 3. (*continued*).



(a) Comparison with baselines



(b) Comparison of ablation studies

Fig. 4. Results on MVTec carpet category. AUC values for c-ss-ms-gVAE with **(a)** baseline methods and **(b)** ablated versions. The plots include bars to show the variability in AUC across randomly sampled training sets, validation sets, and test sets (20 repeats).

reconstruction based loss in the semi-supervised scenario. This is a motivation for c-ss-ms-gVAE to learn not only latent-space distributions but also image-space reconstructions. In addition, all of our frameworks (ss-gVAE, c-ss-gVAE, and c-ss-ms-gVAE) incorporate GG-based models enabling, both, robust heavy-tailed modeling (through the shape parameters) and uncertainty-aware heteroscedastic modeling (through the scale parameters).

The improvement of c-ss-ms-gVAE over other existing baselines (Fig. 4(a)) stem from c-ss-ms-gVAE's modeling and multiscale learning of GG distributions in both latent space and image space; the improvement is significant for $\gamma \geq 0.05$. c-ss-ms-gVAE's improvement over A0 (Fig. 4(b)), especially for $\gamma \geq 0.05$, signifies the importance of the multiscale latent space. c-ss-ms-gVAE's improvement over ss-gVAE (Fig. 4(a)), especially for $\gamma \geq 0.05$, signifies the importance of the multiscale latent space coupled with the latent-space classifier. A0 improves over A1 and A2 (Fig. 4(b)) because of 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 improve significantly, even with

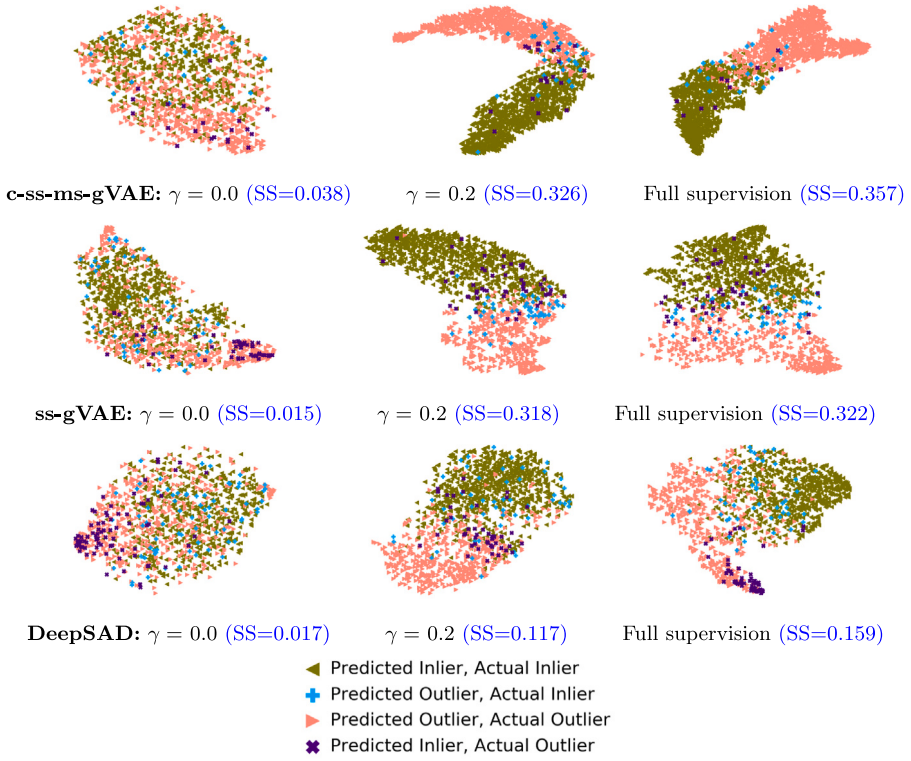


Fig. 5. Results on MVTec's carpet category: t-SNE visualizations of latent-space encodings from the test set for normal and abnormal cells with increasing levels of supervision.

Table 2
Results on 10 MVTec datasets (5 Texture datasets, 5 Object datasets) Comparing 6 methods at $\gamma = 0.2$ supervision. AUC values for c-ss-ms-gVAE with **(a)** baseline methods and **(b)** ablated versions. The plots include bars to show the variability in AUC across randomly sampled training sets, validation sets, and test sets (20 repeats).

Category↓	**c-ss-ms-gVAE**	ss-gVAE	DeepSAD	SSAD	ss-DCAE	BinClass
Carpet	**80.3 ± 1.3**	79.5 ± 2.8	74.3 ± 1.8	72.23 ± 0.2	66.4 ± 2.4	60.0 ± 0.2
Grid	70.3 ± 3.1	71.9 ± 2.4	72.5 ± 2.7	66.86 ± 0.4	59.0 ± 7.9	**76.3 ± 1.2**
Leather	94.3 ± 1.5	92.2 ± 0.3	93.7 ± 0.8	**96.4 ± 0.3**	82.9 ± 1.9	89.4 ± 0.1
Tile	**78.6 ± 2.6**	76.3 ± 1.4	63.6 ± 2.2	74.3 ± 0.1	75.4 ± 1.0	55.8 ± 2.5
Wood	**80.1 ± 1.3**	76.3 ± 3.3	76.2 ± 2.5	67.2 ± 0.1	73.0 ± 1.5	75.1 ± 1.5
Texture: Avg.	**80.7 ± 2.0**	79.2 ± 2.0	76.1 ± 2.0	75.4 ± 0.2	71.3 ± 2.9	71.3 ± 1.1
Bottle	**85.3 ± 3.8**	82.1 ± 4.2	76.9 ± 2.3	52.7 ± 2.0	70.6 ± 5.2	76.2 ± 4.2
Cable	**78.1 ± 0.1**	75.4 ± 1.1	69.2 ± 1.1	57.8 ± 1.1	61.7 ± 0.6	68.9 ± 3.2
Hazelnut	**68.1 ± 2.3**	60.0 ± 1.7	57.2 ± 2.1	53.0 ± 0.6	53.5 ± 2.2	50.3 ± 0.4
Metal nut	**79.3 ± 1.5**	75.5 ± 3.2	71.5 ± 1.5	59.2 ± 0.3	45.0 ± 2.3	74.0 ± 0.2
Screw	**64.7 ± 1.7**	62.3 ± 0.8	61.2 ± 2.5	61.9 ± 0.7	58.0 ± 1.7	59.3 ± 1.5
Object: Avg.	**75.1 ± 1.9**	71.1 ± 2.2	67.2 ± 1.9	56.9 ± 0.9	57.8 ± 2.4	65.7 ± 1.9

little supervision (e.g., $\gamma = 0.05$), over their counterparts that cannot leverage semi-supervision (e.g., DeepSVDD, OC-SVM-Hybrid, ss-DCAE). BinClass, when it becomes applicable at $\gamma > 0$, performs better than some methods, but poorer than DeepSAD, ss-gVAE, and c-ss-ms-gVAE. This is probably because BinClass is prone to overfitting (to the training set) and poor generalization (to the test set) because of the class imbalance during training.

Qualitative Results. For the carpet dataset of MVTec's texture category, Fig. 5 shows the t-SNE visualizations [49] of the latent-space encodings produced/learned by various methods, at different levels of supervision. The t-SNE plots show the separability/overlap between the inliers and outliers. To get a better sense of the separability of the classes in latent space, we use the silhouette score [50] (SS) computed in the latent space for the ground-truth classification. For all methods, as the level of supervision γ increases, SS increases, which is expected. The silhouette scores also indicate that, compared to DeepSAD, both c-ss-ms-gVAE and ss-gVAE (almost) always lead to embeddings that better separate the normal class from the anomalous class.

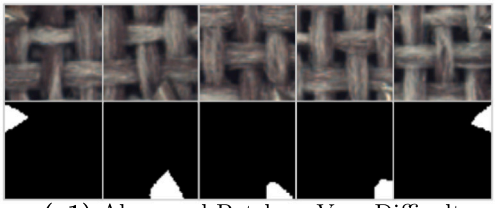
We can go further and evaluate the classification outputs too, as follows. To make this qualitative analysis of classification performance consistent across methods, we can choose the underlying data-classification threshold corresponding to that point in the ROC curve where the sensitivity equals the specificity. The classification outcome is indicated by the labels (i) True Positive (predicted inlier and actual inlier), (ii) True Negative (predicted outlier and actual outlier), (iii) False Positive (predicted inlier and actual outlier), (iv) False Negative (predicted outlier and actual inlier).

Table 3 presents the anomaly scores (described in Section 3.6) for a few samples from carpet category (shown on the left, with their corresponding masks), compared against ss-gVAE and DeepSAD, with and without supervision. For each method-γ combination, we also show the threshold value that is tuned using a validation set, as described in Section 3.6. If the computed anomaly score for the datum is larger than the threshold, then the data point is classified as abnormal, else it is classified as normal.

Table 3(a1) depicts the most challenging (to classify) abnormal-class examples from the carpet category (with the corresponding anomaly mask) which have been misclassified as normal by ss-gVAE and DeepSAD, whereas Table 3(b1) shows some challenging examples that have been correctly classified by c-ss-ms-gVAE and ss-gVAE but misclassified by DeepSAD. In both these cases, the examples belong to the test-set sub-categories that were absent in the training set. Also, the number of pixels labeled abnormal is smaller for the patches in Table 3(a1) as compared to those in Table 3(b1), which makes the former patches more difficult to classify. Table 3(c1) shows some easy to classify patches which have very atypical/abnormal textural characteristics, and are thus correctly classified by all the methods.

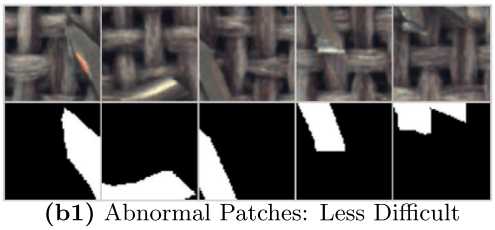
Table 3(a2) shows the most challenging normal-class patches from carpet category which have been correctly classified by c-ss-ms-gVAE but misclassified by ss-gVAE and DeepSAD. These patches tend to have atypical coloration in the textural patterns, unlike the easier patches

Table 3
Anomaly scores across baselines for some of the **(a1)–(c1) abnormal** patches and **(a2)-(b2)** normal patches from MVTec's carpet dataset. The scores in the table are averages across the example patches shown in the adjoining figures. $\gamma = 0$ implies $M = 0$ indicating unsupervised learning; $\gamma = 0.5$ implies $M = N$ indicating fully-supervised learning.



(a1) Abnormal Patches: Very Difficult

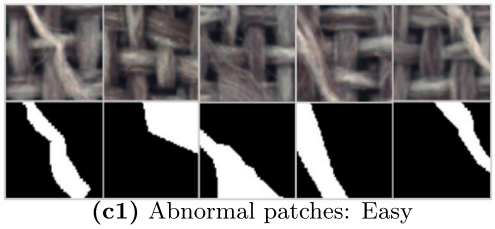
Method	$\gamma = 0$	$\gamma = 0.5$
c-ss-ms-gVAE	**310**	**354**
Threshold	*295*	*329*
ss-gVAE	457	1842
Threshold	*744*	*2128*
DeepSAD	1055	4321
Threshold	*1242*	*4696*



(b1) Abnormal Patches: Less Difficult

Method	$\gamma = 0$	$\gamma = 0.5$
c-ss-ms-gVAE	**348**	**379**
ss-gVAE	**763**	**2251**
DeepSAD	1211	3976

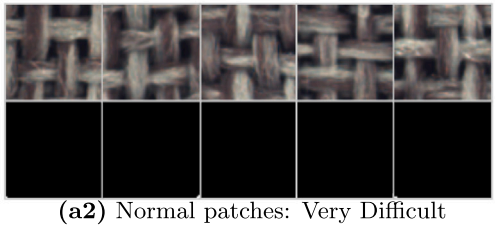
Thresholds same as in (a1)



(c1) Abnormal patches: Easy

Method	$\gamma = 0$	$\gamma = 0.5$
c-ss-ms-gVAE	**314**	**351**
ss-gVAE	**775**	**2942**
DeepSAD	**1284**	**5385**

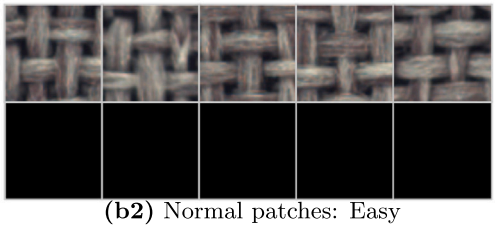
Thresholds same as in (a1)



(a2) Normal patches: Very Difficult

Method	$\gamma = 0$	$\gamma = 0.5$
c-ss-ms-gVAE	**247**	**279**
ss-gVAE	1105	2247
DeepSAD	1429	5431

Thresholds same as in (a1)



(b2) Normal patches: Easy

Method	$\gamma = 0$	$\gamma = 0.5$
c-ss-ms-gVAE	**185**	**315**
ss-gVAE	**689**	**1876**
DeepSAD	**937**	**2488**

Thresholds same as in (a1)

in Table 3(b2) that represent the most typical textural patterns. These visualizations provide insights into the workings of various methods, including ours and the baselines.

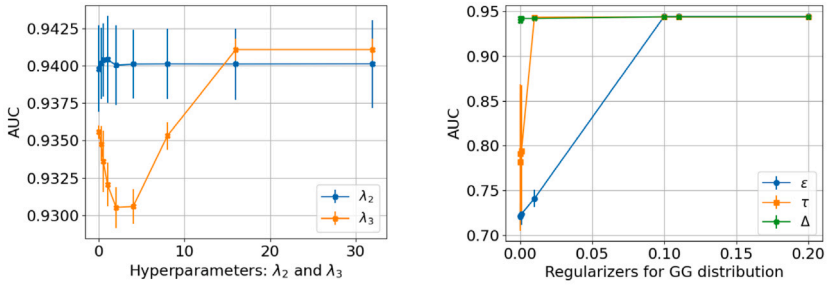
4.4. Robustness analysis for hyperparameter values

We have the following hyperparameters for ss-ms-gVAE: (i) one weighting factor for the two loss terms (same as that in DeepSAD and ss-gVAE), and (ii) regularizers for the generalized-Gaussian distribution to ensure numerical stability, i.e., Δ, ϵ, and τ, which can be all set around 0.15 (similar in spirit to the regularization parameter in DeepSAD). For c-ss-ms-gVAE, there is one extra weighting factor for the additional loss term related to the classifier. Fig. 6 shows the results, on the MVTec's Leather category as an example, of the robustness of the results of c-ss-ms-gVAE with respect to changes in the hyperparameter values. After varying the weighting factors λ_2 and λ_3 over a wide range of values, the results remain very stable; being virtually unchanged for variation over λ_2 and showing only a minor drop in performance for variation over λ_3. Similarly, after varying the regularization parameters for the generalized-Gaussian over a wide range of values, the results remain very stable; i.e., over a roughly 20x variation in the values of τ and Δ, and over a roughly 2x variation in the values of ϵ.

4.5. Results II: Microscopy dataset of malaria-infected blood

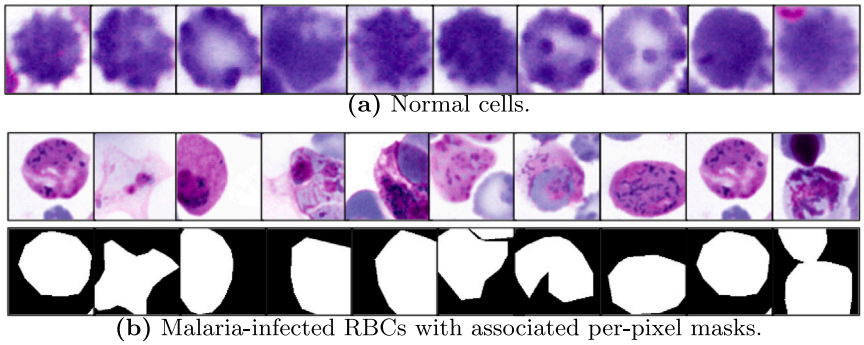
The Broad Bioimage Benchmark Collection [46] (Fig. 7) contains 1328 microscopic images of stained human-blood smears. In this publicly available dataset, each image contains multiple red blood cells (RBCs), some of which are infected with the malarial parasite and are labeled abnormal. Each abnormal RBC has a bounding-box annotation around it. We curated this dataset further by having an expert manually delineate and segment each infected RBC. Because the average cell diameter is around 170 pixels, we perform our analysis on patches of size 170×170 pixels. For empirical analysis, we randomly select 10 000 normal patches and 2500 abnormal patches.

For evaluation of OCC methods, we label a patch as abnormal if at least 50% of its pixels are labeled abnormal. We partition the datasets into three subsets for training, validation, and testing. The validation set is used to tune free parameters. The test set contains a balance of normal and abnormal patches. The images contain two classes of



(a) Weights for loss terms. **(b)** Regularization parameters for GG.

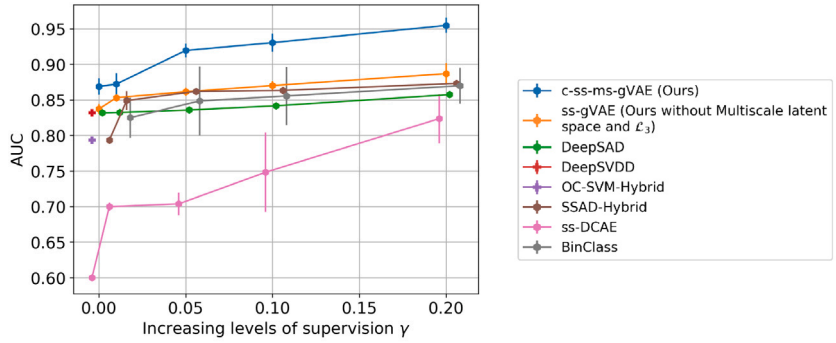
Fig. 6. Results: Robustness analysis for hyperparameter values, using MVTec's Leather category as an example.



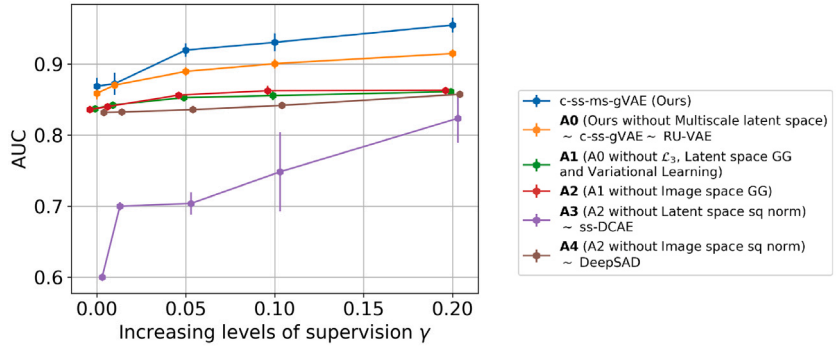
(a) Normal cells.

(b) Malaria-infected RBCs with associated per-pixel masks.

Fig. 7. Microscopy dataset of malaria-infected blood: "Malaria dataset".



(a) Comparison with baselines



(b) Comparison of ablation studies

Fig. 8. Results on Malaria dataset. AUC values for c-ss-ms-gVAE with **(a)** baseline methods and **(b)** ablated versions. The plots include bars to show the variability in AUC across randomly sampled training sets, validation sets, and test sets (20 repeats).

normal cells (RBCs and leukocytes) and four classes of infected cells (gametocytes, rings, trophozoites, and schizonts) [46]. To realize a OCC scenario, out of the available four infected-RBC classes, the training set includes abnormal patches from only two of the infected-RBC classes. The test set uses patches from all six cell classes, thus containing all classes of infected-RBC and normal cells.

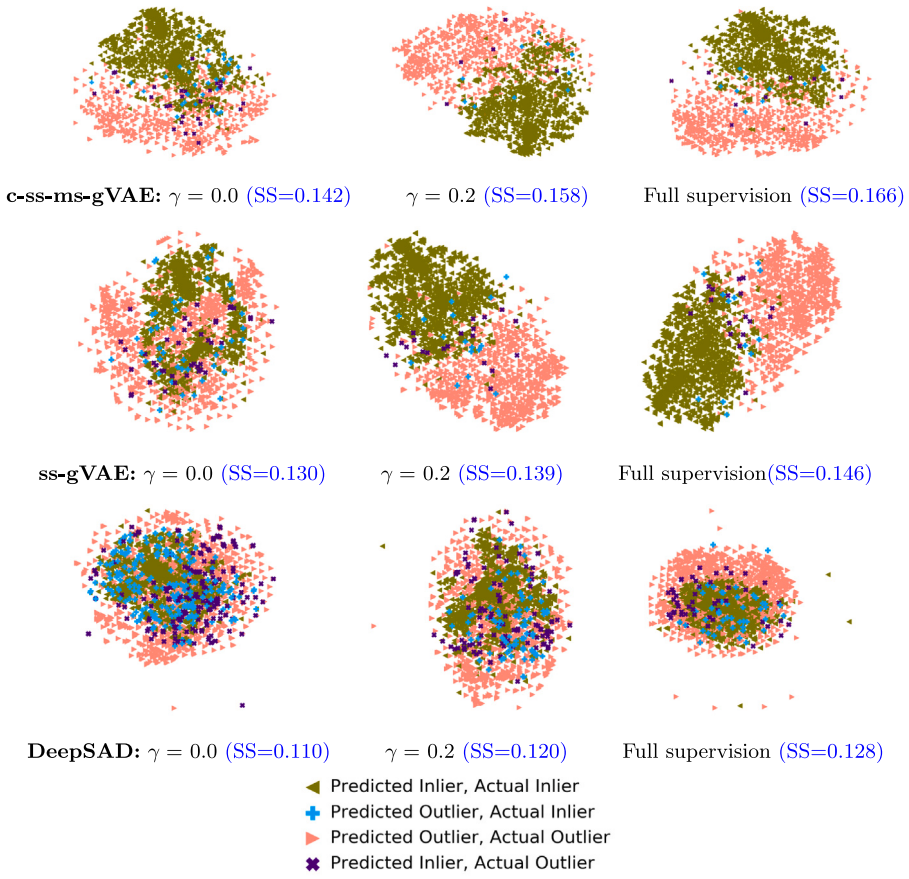


Fig. 9. Results on Malaria dataset: t-SNE visualizations of latent-space distributions from the test set for normal and abnormal cells with increasing levels of supervision.

Fig. 8 shows that, using semi-supervision, DeepSAD improves over its unsupervised version DeepSVDD, and SSAD-Hybrid improves over its unsupervised version OC-SVM-Hybrid, as expected. For this dataset, DeepSAD and DeepSVDD that are solely based on latent-space-loss modeling happen to improve over ss-DCAE that solely models an image-space loss. Nevertheless, the ablated version A0 (equivalent to c-ss-gVAE), which combines GG-based models in latent space and image space improves significantly over (i) both DeepSAD and ss-DCAE (Fig. 8(a)–(b)) as well as (ii) other ablated versions A1–A4 (Fig. 8(b)); this also underscores the importance of both the latent-space and image-space loss terms, along with the use of GG-based variational learning in the latent-space and image-space. Overall, with the further inclusion of a multiscale latent-space as well as a latent-space classifier to the ss-gVAE, our method c-ss-ms-gVAE performs significantly better than all the baselines across all the supervision levels (Fig. 8(a)). BinClass's performance drops more drastically at low levels of semi-supervision γ because of unavailability of enough outlier-class data required for learning a fully-supervised binary classifier; indeed, BinClass cannot perform at $\gamma = 0$.

Qualitative Results. The t-SNE plots of latent-space encodings (Fig. 9) with associated silhouette scores [50] (SS) show improved separability between the normal and abnormal classes for c-ss-ms-gVAE, compared to ss-gVAE and DeepSAD, with unsupervised learning ($\gamma = 0$), with semi-supervised learning ($\gamma = 0.2$), and with full supervision (when $\gamma = 0.5$ or $M = N$). At all supervision levels, c-ss-ms-gVAE has fewer misclassification errors compared to ss-gVAE and DeepSAD. Silhouette score (SS) is a quantitative measure of distinctness of each cluster which is consistent with the t-SNE plots. The SS increases as the amount of supervision increases for the proposed c-ss-ms-gVAE and the baselines. It is evident that c-ss-ms-gVAE and ss-gVAE are able to distinctly separate the normal and anomalous embeddings in Fig. 9 which is consistent with the silhouette scores.

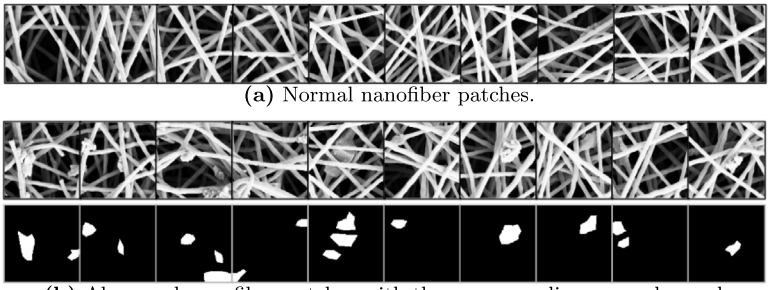
To make this qualitative analysis of classification performance consistent across methods, we choose the underlying data-classification threshold corresponding to that point in the ROC curve where the sensitivity equals the specificity. Our method (c-ss-ms-gVAE) shows improved separability, and improved classification accuracy, at all supervision levels. For instance, at full supervision, the accuracies for c-ss-ms-gVAE, ss-gVAE, and DeepSAD are 78.68%, 44.84%, and 43.92%; at $\gamma = 0.2$, they are 73.36%, 46.2%, and 45%; and at $\gamma = 0$, they are 70.12%, 47.76%, and 44.96%.

4.6. Results III: Electron-microscopy dataset of Nanofiber materials

The Nanofiber dataset [47] (Fig. 10) has 45 electron-microscopy images of nanofiber biomaterials, out of which (i) 5 images are normal and devoid of any anomalies, and (ii) 40 images contain regions with material defects, where the dataset provides the corresponding annotated per-pixel mask for the anomalies. We take patches of size 128 × 128 pixels from these high-resolution images. We consider a patch as abnormal if at least 5% of its pixels are labeled abnormal. We randomly extract 17 228 normal and 5452 abnormal patches.

We partition the datasets into three subsets for training, validation (to tune free parameters), and testing (the test set contains an equal number of normal and abnormal patches). While inputting patches to the DNNs, we resize the patches to 32 × 32. We use the same architecture as in the MVTec dataset and the Malaria dataset, except for a small modification in the first layer to accommodate the gray-scale (single-channel) input in the Nanofiber dataset instead of the RGB (three-channel) input in MVTec and Malaria.

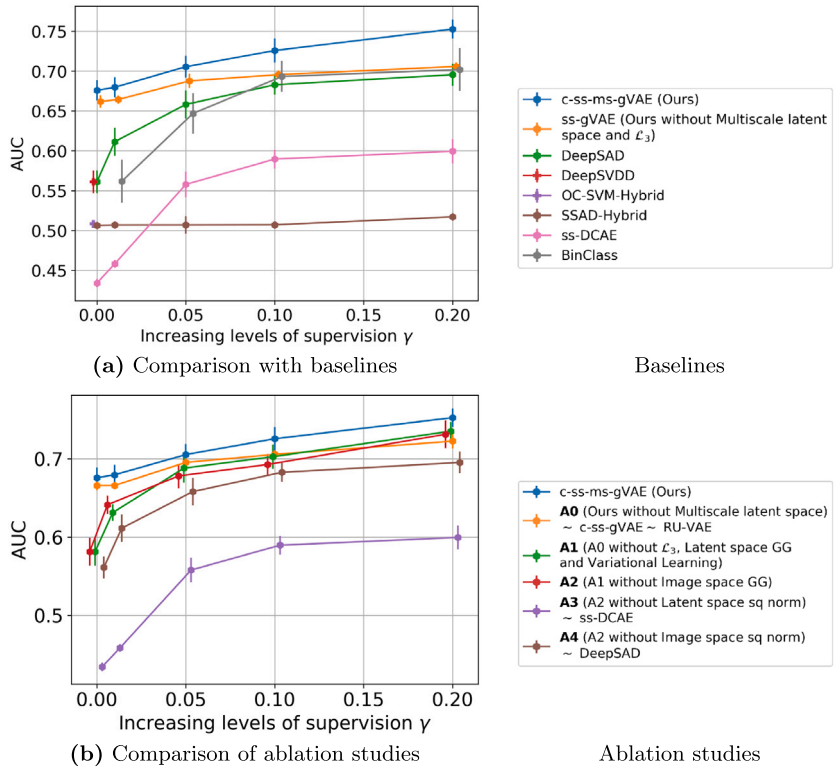
Fig. 11(a) indicates that c-ss-ms-gVAE outperforms the baselines at virtually all levels of supervision. In the unsupervised mode, i.e., when $\gamma = 0$, c-ss-ms-gVAE improve significantly over all other methods.



(a) Normal nanofiber patches.

(b) Abnormal nanofiber patches with the corresponding anomaly mask.

Fig. 10. Electron-microscopy dataset of Nanofiber materials.



(a) Comparison with baselines

(b) Comparison of ablation studies

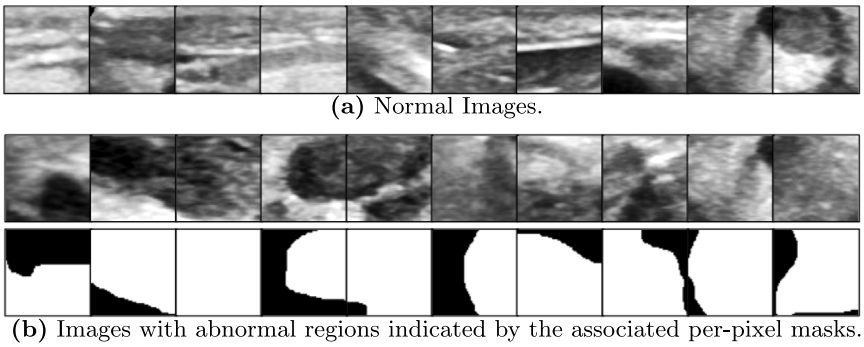
Fig. 11. Results on Nanofiber dataset. AUC values for c-ss-ms-gVAE with **(a)** baseline methods and **(b)** ablated versions. The plots include bars to show the variability in AUC across randomly sampled training sets, validation sets, and test sets (20 repeats).

All semi-supervised methods show significant improvements over their unsupervised counterparts in the case of Nanofibre dataset as well. The ablation studies in Fig. 11(b) underscore the importance of combining latent-space and image-space modeling (A2 improves significantly over A3 and A4) and the multiscale latent space (c-ss-ms-gVAE improves over A0) within the proposed framework. BinClass's performance drops more drastically at low levels of semi-supervision γ because of the class imbalance in the training set.

4.7. Results IV: Breast ultrasound image dataset

The fourth dataset (Fig. 12) comprises breast ultrasound images [48], collected in the year 2018, for 600 patients. It consists of 780 images with a typical image size of 500 × 500 pixels. The images are categorized into three classes, i.e., normal, benign, and malignant. Ground-truth per-pixel segmentation of the anomalies are available for each image. We take patches of size 64 × 64 because, the anomalous regions are typically of sizes around 64 × 64 pixels. For DNN training, we resize the patches to 32 × 32 pixels, and use the same architecture as that used for the Nanofiber dataset.

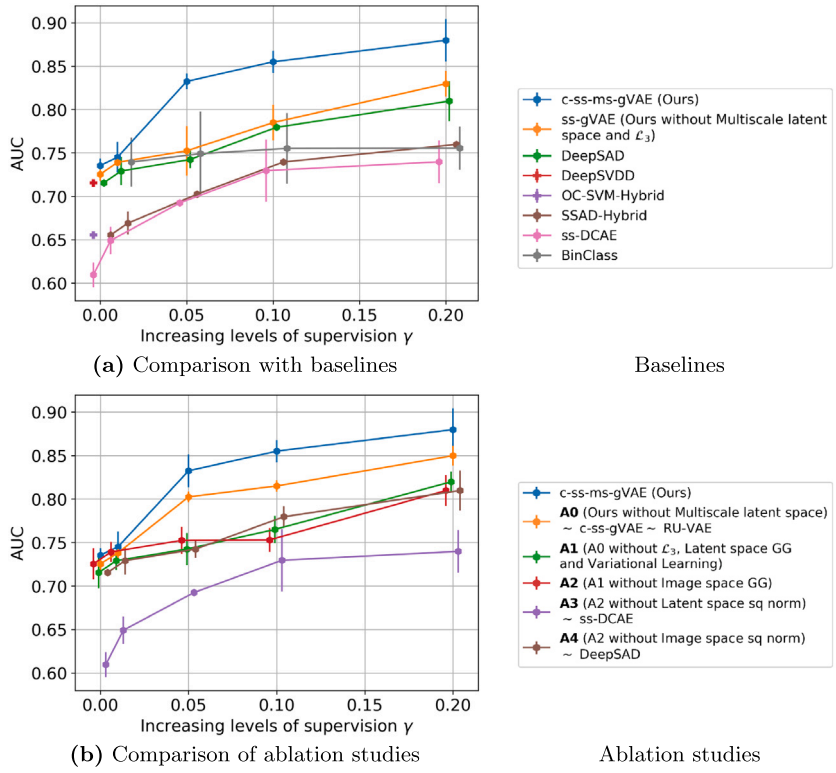
For training and validation, we take patches from the normal class and the benign class (as abnormal). Consistent with a typical OCC scenario, the malignant-class (abnormal) patches appear only in the test set. A patch is considered abnormal if at least 50% of its pixels contain an anomaly (as indicated by the corresponding ground-truth mask). For empirical analysis, we randomly select 16 000 patches for the training set and 5000 patches for the test set. Each set has a uniform balance of normal and abnormal patches in it. Fig. 13 shows that c-ss-ms-gVAE performs significantly better than the baselines and its ablated versions in the presence of supervision (e.g., $\gamma \geq 0.05$) during training.



(a) Normal Images.

(b) Images with abnormal regions indicated by the associated per-pixel masks.

Fig. 12. Breast ultrasound images dataset.



(a) Comparison with baselines

(b) Comparison of ablation studies

Fig. 13. Results on breast ultrasound dataset. AUC values for c-ss-ms-gVAE with **(a)** baseline methods and **(b)** ablated versions. The plots include bars to show the variability in AUC across randomly sampled training sets, validation sets, and test sets (20 repeats).

Consistent with the analyses for the other datasets, semi-supervised methods for anomaly detection perform better than their unsupervised counterparts (Fig. 13(a)). Also, the ablation studies in Fig. 13(b) signify the importance of GG modeling because A0 outperforms the ablated versions A1–A4 in the presence of some supervision. The incorporation of a multiscale latent space in c-ss-ms-gVAE also shows a significant improvement in performance compared to A0.

5. Conclusion

This paper proposes a novel generalized-VAE theoretical framework for statistical learning (c-ss-ms-gVAE) incorporating (i) generalized-Gaussian modeling in both latent space and image space, (ii) a multiscale latent space promoting a rich feature set at multiple spatial scales/resolutions, and (iii) semi-supervised learning that enables it to leverage a small subset of labeled data during training for OCC. c-ss-ms-gVAE significantly extends our preliminary work ss-gVAE by (i) designing a multiscale latent-space architecture within the theoretical framework and, (ii) during semi-supervised learning, incorporating a non-linear classifier to discriminate between the manifolds of latent-space encodings of the inlier and outlier classes. The generalized-Gaussian distributions model datum-specific heavy-tailed distributions (robust statistical modeling) and heteroscedastic (uncertainty-aware) modeling in the image space and latent space. To enable backpropagation based optimization with the generalized-Gaussian modeling, c-ss-ms-gVAE proposes a reparameterization scheme relying on Gamma random variables and implicit reparameterization. Comprehensive quantitative and qualitative empirical analyses (using six baselines, five ablated

versions, ten industrial datasets, and three medical datasets), provide insights into the benefits of c-ss-ms-gVAE over existing methods.

The paper has analyzed carefully designed ablated versions to identify the importance of each component underlying our proposed framework and to demonstrate its competence when compared with methods which lack those components. During model learning of our method, we employ sequential warm-start based optimization for efficacy, as described in Section 3.7. The proposed framework has real-world implications in manufacturing, medical, and transport industries to identify defects in the images at patch levels. The uncertainty estimates from our method can go a long way in improving the human interpretation of the results as well as informing the subsequent stages in automated data processing. The theoretical framework underlying our method is very general and can lend itself to various real-world tasks and application scenarios. Our proposed method, as with many variational DNN methods, is computationally intensive during training, and thus it is our endeavor is to make computationally light-weight versions of it customized for different applications; such lighter-weight versions may arise from modified DNN architectures. At the same time, future work will analyze the practical feasibility of running the learned models at test times on portable computing devices.

CRediT authorship contribution statement

Renuka Sharma: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Suyash P. Awate:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Conceptualization, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is publicly available and has been cited in the manuscript.

References

[1] S. Chen, Y. Liu, C. Liu, T. Chen, Y. Wang, Domain-generalized textured surface anomaly detection, in: Int. Conf. on Mul. and Exp., 2022, pp. 01–06.
[2] L. Nie, L. Zhao, K. Li, Glad: Global and local anomaly detection, in: Int. Conf. on Mul. and Exp., 2020, pp. 1–6.
[3] R. Sharma, H. Shi, J. Cai, S.P. Awate, N. Birbilis, Deep semi-supervised anomaly detection using VQ-VAE, in: Int. Conf. on Dig. Image Comp.: Tech.s and App., 2023, pp. 273–280.
[4] S. Marimont, G. Tarroni, Anomaly detection through latent space restoration using vector quantized variational autoencoders, in: Int. Symp. on Bio. Ima., 2021, pp. 1764–1767.
[5] M. Kimura, T. Yanagihara, Anomaly detection using GANs for visual inspection in noisy training data, in: Asi. Conf. on Comp. Vis, 2018, pp. 373–385.
[6] C. Ma, Z. Miao, M. Li, S. Song, M. Yang, Detecting anomalous trajectories via recurrent neural networks, in: Asi. Conf. on Comp. Vis, 2018, pp. 370–382.
[7] H.L. Kennedy, Whitening pre-filters with circular symmetry for anomaly detection in hyperspectral imagery, in: Dig. Ima. Comp.: Tech. and App., DICTA, IEEE, 2018, pp. 1–8.
[8] R. Kommanduri, M. Ghorai, DAST-Net: Dense visual attention augmented spatio-temporal network for unsupervised video anomaly detection, Neurocomputing 579 (2024) 127444.
[9] Z. Wang, X. Gu, J. Hu, X. Gu, Ensemble anomaly score for video anomaly detection using denoise diffusion model and motion filters, Neurocomputing (2023) 126589.
[10] W. Luo, W. Liu, D. Lian, S. Gao, Future frame prediction network for video anomaly detection, IEEE Trans. Pattern Anal. Mach. Intell. (2021) 7505–7520.
[11] Y. Xia, X. Cao, F. Wen, G. Hua, J. Sun, Learning discriminative reconstructions for unsupervised outlier removal, in: ICCV, 2015, pp. 1511–1519.
[12] D. Kingma, M. Welling, Auto-encoding variational Bayes, in: ICLR, 2014.
[13] L. Ruff, et al., Deep one-class classification, in: ICML, 2018, pp. 4393–4402.
[14] L. Ruff, et al., Deep semi-supervised anomaly detection, in: ICLR, 2020.
[15] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: Int Conf on Med Ima Comp and Comp-Assisted Int, 2015, pp. 234–241.
[16] R. Sharma, S. Mashkaria, S.P. Awate, A semi-supervised generalized VAE framework for abnormality detection using one-class classification, in: Winter Conf on App of Comp Vis, WACV, 2022, pp. 595–603.
[17] H. Hoffmann, Kernel PCA for novelty detection, Pattern Recognit. 40 (3) (2007) 863–874.
[18] B. Schölkopf, R. Williamson, A. Smola, J. Shawe-Taylor, J. Platt, Support vector method for novelty detection, in: NIPS, 2000, pp. 582–588.
[19] D. Tax, R. Duin, Support vector data description, Mach. Learn. 54 (1) (2004) 45–66.
[20] R. Campello, D. Moulavi, A. Zimek, J. Sander, Hierarchical density estimates for data clustering, visualization, and outlier detection, ACM Trans. Knowl. Discov. Data 10 (1) (2015) 1–51.
[21] H. Hojjati, N. Armanfard, Dasvdd: Deep autoencoding support vector data descriptor for anomaly detection, IEEE Trans. Knowl. Data Eng. (2023).
[22] T. Schlegl, P. Seeböck, S.M. Waldstein, U. Schmidt-Erfurth, G. Langs, Unsupervised anomaly detection with generative adversarial networks to guide marker discovery, in: IPMI, 2017, pp. 146–157.
[23] J. Masci, U. Meier, D. Cireşan, J. Schmidhuber, Stacked convolutional auto-encoders for hierarchical feature extraction, in: ICANN, 2011, pp. 52–59.
[24] R. Chalapathy, A. Menon, S. Chawla, Robust, deep and inductive anomaly detection, in: ECML PKDD, 2017, pp. 36–51.
[25] C. Zhou, R. Paffenroth, Anomaly detection with robust deep autoencoders, in: SIGKDD, 2017, pp. 665–674.
[26] W. Liu, G. Hua, J. Smith, Unsupervised one-class learning for automatic outlier removal, in: CVPR, 2014.
[27] G. Kim, J.G. Choi, M. Ku, S. Lim, Developing a semi-supervised learning and ordinal classification framework for quality level prediction in manufacturing, Comput. Ind. Eng. 181 (2023) 109286.
[28] G. Kim, S.M. Yang, D.M. Kim, S. Kim, J.G. Choi, M. Ku, S. Lim, H.W. Park, Bayesian-based uncertainty-aware tool-wear prediction model in end-milling process of titanium alloy, Appl. Soft Comput. 148 (2023) 110922.
[29] S. Akcay, A. Atapour-Abarghouei, T.P. Breckon, GANomaly: Semi-supervised anomaly detection via adversarial training, in: Computer Vision – ACCV 2018, Springer International Publishing, 2019, pp. 622–637.
[30] R. Sharma, S.P. Awate, Robust and uncertainty-aware VAE (RU-VAE) for one-class classification, in: Int Symp on Bio Ima, ISBI, 2022, pp. 1–5.
[31] D.-H. Lee, et al., Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks, in: Workshop on Chall in Rep Lear, ICML, 2013, p. 896.
[32] A. Blum, T. Mitchell, Combining labeled and unlabeled data with co-training, in: Conf on Comp Lear Th, 1998, pp. 92–100.
[33] S. Melacci, M. Belkin, Laplacian support vector machines trained in the primal, J. Mach. Learn. Res. 12 (Mar) (2011) 1149–1184.
[34] E. Bauman, K. Bauman, One-class semi-supervised learning, in: Braverman Readings in Machine Learning. Key Ideas from Inception to Current State, 2017, pp. 189–200.
[35] N. Gornitz, M. Kloft, K. Rieck, U. Brefeld, Toward supervised anomaly detection, J. Artificial Intelligence Res. 46 (1) (2013) 235–262.
[36] M.P. Shah, S.N. Merchant, S.P. Awate, Abnormality detection using deep neural networks with robust quasi-norm autoencoding and semi-supervised learning, in: ISBI, 2018, pp. 568–572.
[37] A. Kendall, Y. Gal, What uncertainties do we need in Bayesian deep learning for computer vision? in: NIPS, 2017, pp. 5574–5584.
[38] C. Leibig, V. Allken, M.S. Ayhan, P. Berens, S. Wahl, Leveraging uncertainty information from deep neural networks for disease detection, Sci. Rep. 7 (1) (2017) 1–14.
[39] J.M. Dolezal, A. Srisuwananukorn, D. Karpeyev, S. Ramesh, S. Kochanny, B. Cody, A.S. Mansfield, S. Rakshit, R. Bansal, M.C. Bois, et al., Uncertainty-informed deep learning models enable high-confidence predictions for digital histopathology, Nat. Commun. 13 (1) (2022) 6572.
[40] M. Nardon, P. Pianca, Simulation techniques for generalized Gaussian densities, J. Stat. Comput. Simul. 79 (11) (2009) 1317–1329.
[41] M. Figurnov, S. Mohamed, A. Mnih, Implicit reparameterization gradients, in: NIPS, Vol. 31, 2018, pp. 441–452.
[42] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, in: ICLR 2015.
[43] V.L. Cao, M. Nicolau, J. McDermott, A hybrid autoencoder and density estimation model for anomaly detection, in: PPSN, 2016, pp. 717–726.
[44] S.M. Erfani, S. Rajasegarar, S. Karunasekera, C. Leckie, High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning, Pattern Recognit. 58 (2016) 121–134.
[45] P. Bergmann, M. Fauser, D. Sattlegger, C. Steger, MVTec AD–A comprehensive real-world dataset for unsupervised anomaly detection, in: CVPR, 2019, pp. 9592–9600.

[46] V. Ljosa, K. Sokolnicki, A. Carpenter, Annotated high-throughput microscopy image sets for validation, Nature Methods 9 (2012) 637.
[47] D. Carrera, F. Manganini, G. Boracchi, E. Lanzarone, Nanofibre dataset, 2021, http://www.mi.imati.cnr.it/ettore/NanoTWICE/, [Online; Accessed 26 February 2021].
[48] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images, Data Brief 28 (2020) 104863.
[49] L. V. Maaten, G. H, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).
[50] K.R. Shahapure, C. Nicholas, Cluster quality analysis using silhouette score, in: Int Conf on Data Sc and Adv Analy, DSAA, 2020, pp. 747–748.

Renuka Sharma is a postdoctoral fellow at Commonwealth Scientific and Industrial Research Organisation, Australia. She holds a joint-PhD from Data Science and AI Department at Monash University, Melbourne and Department of Computer Science and Engineering at the Indian Institute of Technology Bombay, Mumbai, India. Her research Interests include machine learning, computer vision, multi-modal learning, and embedded AI. More details are available at https://renukasharma.github.io/.

Suyash P. Awate is the Asha and Keshav Bhide Chair Professor in the Department of Computer Science and Engineering at the Indian Institute of Technology Bombay, Mumbai, India. His research Interests include image analysis, machine learning, medical image computing, and computer vision. More details are available at https://www.cse.iitb.ac.in/~suyash/.