Towards Open-World Object-based Anomaly Detection via Self-Supervised Outlier Synthesis

Brian K. S. Isaac-Medina*, Yona Falinie A. Gaus*,
Neelanjan Bhowmik*, Toby P. Breckon*,†

Department of {*Computer Science, †Engineering}, Durham University, UK

Abstract. Object detection is a pivotal task in computer vision that has received significant attention in previous years. Nonetheless, the capability of a detector to localise objects out of the training distribution remains unexplored. Whilst recent approaches in object-level out-of-distribution (OoD) detection heavily rely on class labels, such approaches contradict truly open-world scenarios where the class distribution is often unknown. In this context, anomaly detection focuses on detecting unseen instances rather than classifying detections as OoD. This work aims to bridge this gap by leveraging an open-world object detector and an OoD detector via virtual outlier synthesis. This is achieved by using the detector backbone features to first learn object pseudo-classes via self-supervision. These pseudo-classes serve as the basis for class-conditional virtual outlier sampling of anomalous features that are classified by an OoD head. Our approach empowers our overall object detector architecture to learn anomaly-aware feature representations without relying on class labels, hence enabling truly open-world object anomaly detection. Empirical validation of our approach demonstrates its effectiveness across diverse datasets encompassing various imaging modalities (visible, infrared, and X-ray). Moreover, our method establishes state-of-the-art performance on object-level anomaly detection, achieving an average recall score improvement of over 5.4% for natural images and 23.5% for a security X-ray dataset compared to the current approaches. In addition, our method detects anomalies in datasets where current approaches fail. Code available at `https://github.com/KostadinovShalon/oln-ssos`.

1 Introduction

Anomaly detection plays a crucial role in identifying deviations from the norm in various applications such as industrial inspection [12,30,47] and video surveillance [17,32,37,38]. In general, anomaly detection addresses an aspect of the open set problem in computer vision - whilst normality in terms of the appearance and behaviour of objects within the scene can be bounded, conversely, the set of possible anomalous occurrences is unbounded. Anomalous events rarely occur as compared to normal activities, which in itself results in the commonplace dataset challenge of anomaly detection - the availability of *abnormal* (anomalous) samples is limited in both volume and variety. This in itself leads to a naturally imbalanced dataset distribution for any real-world anomaly detection problem. A common approach is to learn a model of the normal (non-anomalous) data

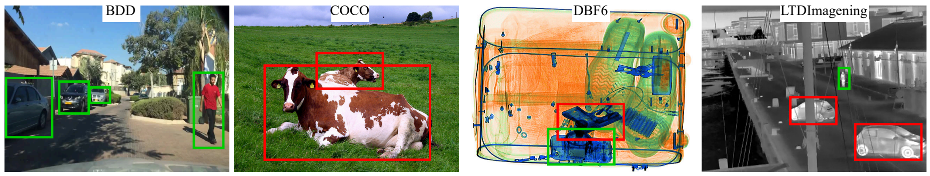

Fig. 1: Exemplar of in-distribution (green) and out-of-distribution (red) objects from across four diverse benchmark datasets with various imaging modalities.

distribution from the abundance of normal sample training data available and then detect anomalies as outliers in a semi-supervised manner [2, 3, 47].

In this context, out-of-distribution (OoD) detection identifies instances from unknown (unseen) classes by training only on in-distribution data. For instance, works of [28, 34] approach OoD detection by measuring the joint probability of a sample coming from one of the training classes using the free energy of the sample, subsequently identifying never-seen-before (outlier) objects. Whilst this approach has proven effective, its application within the object detection framework faces the challenge of localising unseen object categories. For example, a self-driving vehicle may encounter wild animals [15], an X-ray security screening system may identify prohibited items [2] or unknown vehicles in surveillance systems [17] (Fig. 1). These anomalous variations can vary from the visually obvious to the very subtle in the object [3] and, while current OoD approaches can classify them as anomalies, the detector must first localise them within the image. This challenge serves as motivation for this work, differentiating object-based anomaly detection from out-of-distribution detection by defining it as the joint task of localising unseen objects and identifying them as anomalous.

This work proposes an object-based anomaly detection framework that leverages an open-world object detector (OWOD), which localises unseen objects without prior class supervision. Whilst classic object detectors [6, 46, 61] are focused on detecting objects from a set of known categories, open world object detectors [20, 26, 73] naturally capture both known and unknown objects. However, the lack of classification in such OWOD subsequently disables the use of secondary class-based OoD detectors. To overcome this challenge, our method learns object pseudo-labels in a self-supervised manner consisting of alternating deep feature clustering and neural network-based prediction similar to DeepCluster [7]. We cluster the object features from the detector backbone, pooled by a RoIAlign layer [21], before each training epoch such that the resulting assignments from the clustering are used as ground truth labels to train a classifier head. Following the virtual outlier synthesis (VOS) methodology [15], class-conditional Gaussian distributions are subsequently learnt in the object feature space, which are used to sample virtual outliers from low-likelihood regions. Both in-distribution objects and virtual outliers are finally used to train a normal vs. abnormal classifier head. Since virtual outliers are sampled from the pseudo-label object distribution, we call this approach as Self-Supervised

Outlier Synthesis (SSOS). To the best of our knowledge, we are the first to use self-supervised pseudo-classification on object instances for energy-based OoD detection, enabling object-level class-agnostic anomaly detection. As we use the object localisation network (OLN) [26] as our OWOD, we denote our method as **OLN-SSOS**. We evaluate over a diverse set of imaging modalities and applications, successfully detecting anomalous instances across a wide variety of contexts. In summary, our main contributions are:

- **the first class-agnostic, end-to-end object-based anomaly detection architecture** that learns object pseudo-labels to fit class-conditional Gaussian distributions in the object feature space of an OWOD, thus enabling energy-based OoD detection; our approach uses self-supervised outlier synthesis (SSOS) to identify anomalies not in the training set.
- **state-of-the-art anomaly detection performance across four diverse benchmark datasets**, namely BDD100K/COCO [68], LTDImaging [42], SIXRay10 [39] and DBF6 [1], and competitive performance on the VOC/COCO [16, 33] benchmark. Our method achieves an average recall improvement of 5.4% for the BDD100k/COCO, 23.5% for DBF6, and establishes the state-of-the-art for SIXRay10 and LTDImaging, where current OoD approaches fail. In addition, it achieves an average recall of 17.8% (vs. 20.6% using VOS [15]) without class supervision.
- **qualitative analysis** illustrating that our architecture can jointly localise previously unseen objects within an image and classify them as anomalous, whilst other methods only identify OoD objects that are similar to the in-distribution dataset and can hence be localised by the class-based object detector (*e.g.*, animals and vehicles are present in the training and test datasets as super-classes, with a subset of intra-class anomalous instances therein).
- **supporting ablation studies** illustrating the impact of our methodological design choices, such as the number of clusters or the use of instance masks.

2 Literature Review

The terminology of anomaly detection, outlier detection and out-of-distribution (OoD) detection are largely used interchangeably in the literature to describe tasks whereby the primary goal is to model the norm of a given problem domain and hence detect (outlier) deviations from that model. In general, anomaly (or outlier) detection (Sec. 2.1) operates under the assumption that data sample availability is highly biased towards normal classes whilst inadequate distribution coverage exists for other (abnormal) classes which may be unbounded in nature [17]. By contrast, OoD detection (Sec. 2.2) leverages a closed set problem whereby in-distribution samples belong to one of a predefined set of class labels, and hence outliers are objects that do not fit into any of those category labels.

2.1 Anomaly Detection

The approaches for detecting anomalies in images fall into three main categories: feature embedding, reconstruction-based, and streaming-based approaches.

Within feature embedding techniques, well-known methods include memory bank [10,11,30,47], knowledge distillation [4,12,53], normalising flow networks [19,50,69], and one-class classification strategies [51,52,67,70]. Among these, the memory bank approach, exemplified by PatchCore [47], followed by SPADE [10], Padim [11], and CFA [30], stands out for their effectiveness. These methods extract features from all normal images and store them in a memory bank. During testing, image features are matched against the normal features stored in the memory. However, the effectiveness of these methods heavily relies on the completeness of the memory, which requires a vast collection of normal images to fully capture the normal pattern. Additionally, the size of the memory is often tied to the size of the dataset or the dimensions of the images, rendering these approaches impractical for scenarios involving large or high-resolution datasets.

Reconstruction-based methods address anomaly detection at pixel level by typically employing autoencoders [5,36,56] or generators [2,66,71] to encode and decode the input normal images, indirectly learning the distribution of normal images through the process of reconstruction. While these algorithms deliver good results, they encounter difficulties with objects of complex textures and structures. Consequently, they are prone to reconstruction errors, compromising their ability to differentiate between normal and anomalous instances.

Temporal streaming-based approaches are mostly applied in video clips [24, 35, 40, 41, 43, 52] where the task primarily involves detecting unusual events or behaviours within normal events. This is primarily achieved by analysing object trajectories [31, 48, 49] and motion characteristics [35, 41, 45, 62]. For instance, Roy et al. [48, 49] incorporate deep autoencoders to model the trajectories of normal events, subsequently identifying any abnormal trajectories as outliers. On the other hand, the work of [35] introduces an optical flow loss as a motion constraint during training. In contrast, the works of [41, 45] focus on learning motion by predicting the optical flow of the current frame. Additionally, the works of [17, 62] use optical flow information to guide frame prediction, where motion knowledge is used to discriminate between normal and abnormal frames.

2.2 OoD Detection

Earlier works on out-of-distribution detection (OoD) focus on the use of generative models as a means to model the in-distribution data classes [18, 29, 57, 60]. Whilst such methods give satisfactory performance, this often drops off with increased dataset diversity and image fidelity resulting in the more recent advent of feature-based synthetic outlier generation techniques in the OoD space.

Du et al. [15] introduce virtual outlier synthesis (VOS) for OoD, employing inlier features to fit class-conditional Gaussian distributions and sampling OoD features from low likelihood regions of these distributions. Under a similar approach, [14]incorporate unknown-aware knowledge from auxiliary videos to effectively improve the performance of distinguishing OoD objects. Kumar et al. [28] argue that synthesising outlier features from class-wise low-likelihood regions does not ensure that these features will not overlap other class high likelihood regions. Therefore, they use an invertible normalising flow taking all in-distribution objects into a common feature space where outliers are sampled, improving over the decision boundary between in-distribution and OoD objects.

Whilst prior work [14, 15, 28] concentrates on feature-based OoD object detectors, [63] leverages the backbone of an object detector network, identifying that residual convolutional layers with batch normalisation are the most effective layers for identifying OoD samples. Another notable work [13] slightly deviates from this previous approach of using class-conditional Gaussian distributions [14, 15, 28], by utilizing von Mises-Fisher (vMF) distributions to shape the learned representation for detecting OoD objects.

Similar to our approach, the works of [20, 25, 26, 72] use OWOD to identify both known and unknown classes by training on pseudo-labelled unknown objects while continuously acquiring updated annotations for new unseen classes. For instance, Gupta et al. [20] introduce a Transformer-based framework with multi-scale self-attention to discriminate between (open-set) objects and background. Wu et al. [65] incorporates a two-stage object detector to classify objects into different unknown classes, while Zhao et al. [72] use a more traditional approach (selective search) to correct the auto-labelled first-stage region-proposals and subsequently classify unknown instances into new classes.

Whilst these approaches exhibit good unknown object detection performance to subsequently identify anomaly/OoD occurrences, most of the aforementioned methods present several notable challenges. First, detecting abnormalities rely heavily on suitable access to real outlier data samples [32,37,38,58], or a complex generative process to synthesise such samples [24, 35, 40, 41, 43, 52]. In real-world scenarios, anomalies can vary from the visually obvious (e.g. person dressed as a clown) [8] to the very subtle (e.g. descending fog or mist due to adverse weather condition) [42]. Second, all the aforementioned work explicitly relies upon existing object-wise class labels in order to detect out-of-distribution occurrences [13–15, 28]. This contrasts sharply with reality, where unknown (unlabelled) object classes will naturally occur and anomaly occurrences will be a rarity. As a result, incorporating a class-agnostic OWOD is a crucial step towards building reliable object-wise anomaly detection for real-world scenarios.

3 Methodology

Our method consists of the combination of three architectures for different tasks that jointly enable class-agnostic object-based anomaly detection. First, an OWOD is used to detect all possible objects within the scene (Sec. 3.1); subsequently, an unsupervised classifier head learns to cluster object categories by pooling features from the backbone of the OWOD (Sec. 3.2); and finally, a self-supervised outlier synthesis module produces a set of virtual outliers by sampling low-likelihood regions from the feature space using the learned pseudo-labels (Sec. 3.3). The overall architecture is illustrated in Fig. 2.

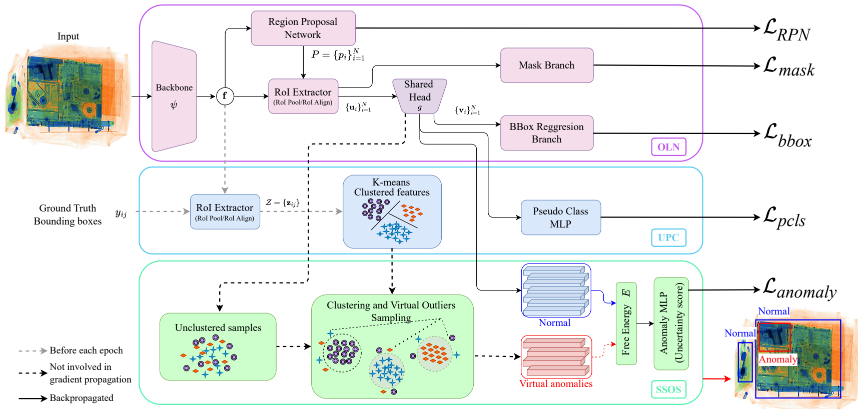

Fig. 2: Our proposed architecture for open-world anomaly detection combines the object localisation network (OLN), unsupervised pseudo-classification (UPC) and anomaly detection via self-supervised outlier synthesis (SSOS).

3.1 Class-agnostic Open-world Object Detection

In order to detect anomalies, a detector capable of localising objects belonging to unknown classes not available during training is needed. To this end, we adopt the object localisation network (OLN) architecture [26] to predict the maximal number of objects within an image. Different from standard detection, the OWOD localise *all* possible objects in an image in a class-agnostic manner.

As depicted in Fig. 2, the OLN consists of a region proposal network (RPN) [46], a bounding box regression head and an optional mask branch. In this context, given an input image $\mathbf{x}$, a 2D feature map $\mathbf{f} = \psi(\mathbf{x})$ is extracted by a backbone network ψ. Subsequently, the RPN predicts a set of N proposal bounding boxes $P = \{p_i\}_{i=1}^N, p_i \in \mathbb{R}^4$. Each object candidate p_i is parameterised by a point, usually the top-left corner, and the bounding box width and height. In a departure from the classical RPN design introduced in Faster R-CNN [46] that classifies region proposals as foreground or background, the RPN in OLN regresses the centreness [59] c_i of the bounding box with the maximal overlapping ground truth bounding box as a measurement of proposal quality. Additionally, the parameters of valid proposals, *i.e.*, those with an intersection over union (IoU) greater than a threshold, are also regressed. L1 Losses are used for the centreness and the box parameters. The RPN loss is:

$$\mathcal{L}_{RPN} = \frac{1}{N}\sum_{i}^{N} \text{L1Loss}(c_i, \hat{c}_i) + \mathbb{1}_{obj}\text{L1Loss}(p_i, \hat{p}_i)\,, \tag{1}$$

where $\mathbb{1}_{obj}$ is 1 if the proposal is matched against a ground truth (0 otherwise). Next, the proposal features $\mathbf{u}_i$ are extracted from $\mathbf{f}$ using RoIAlign [21]. These features are subsequently fed into a shared network g (later used at different stages) producing an object representation $\mathbf{v}_i = g(\mathbf{u}_i) \in \mathbb{R}^d$. Similarly to the

RPN, a bounding box branch regresses the box quality b_i and the bounding box parameters p using L1 losses. The bounding box loss is:

$$\mathcal{L}_{bbox} = \frac{1}{N_v} \sum_{i}^{N_v} \text{L1Loss}(b_i, \hat{b}_i) + \text{L1Loss}(p_i, \hat{p}_i) \,, \tag{2}$$

where N_v is the number of valid proposals and the box quality b_i is given by the IoU with the ground truth. An optional mask branch is added, consisting of the same mask head as in Mask R-CNN [21] and a mask scoring network [23] that predicts a mask quality score m_i. The mask loss $\mathcal{L}_{mask}$ is hence the sum of the Mask R-CNN and Mask Scoring mask losses. During inference, the prediction score is calculated as $s_i = \sqrt{c_i b_i}$, or $s_i = \sqrt[3]{c_i b_i m_i}$ if the mask branch is used.

3.2 Unsupervised Pseudo Classification

As it will be discussed in Sec. 3.3, OoD driven object-wise anomaly detection relies on the categorical distribution of the ground truth object classes at training time. However, given that our open-world object detector is class-agnostic, ground truth class labels are not available. For this reason, we perform Unsupervised Pseudo Classification (UPC) following the DeepCluster [7] methodology of learning pseudo-classes from underlying deep feature representations. In this regard, the UPC strategy consists of clustering a set of M feature vectors $f_i \in \mathbb{R}^d$ into k classes, using the resulting pseudo-labels to train a classifier network. Clustering is performed several times during the training process such that the clusters are re-assigned to accommodate more recently learned representations.

While DeepCluster is implemented for image classification networks, we implement UPC by clustering object features from the backbone using a RoIAlign layer. Feature clustering is performed before each epoch via K-means trained on ground truth objects. Let $\mathcal{Z} = \{\mathbf{z}_{ij}\}, \mathbf{z}_{ij} = \text{RoIAlign}(\psi(\mathbf{x}_i), y_{ij})$ be the set of the feature representations of all ground truth bounding boxes y_{ij} from all training images $\mathbf{x}_i$. K-means clustering is performed on $\mathcal{Z}$ into K clusters, producing a set of cluster centres $\mathbf{w}_1, \ldots, \mathbf{w}_K$. Subsequently, each ground truth bounding box is assigned a *pseudo-label* after each epoch t such that:

$$l_{ij}^{(t)} = \arg\min_k d(\mathbf{z}_{ij}, \mathbf{w}_k) \,, \tag{3}$$

where $l_{ij}^{(t)}$ is the bounding box b_{ij} label after t epochs and d is the L2 distance. Finally, a multi-layer perception (MLP) predicts the pseudo-class logits f_k for each object and Cross Entropy is used as the pseudo-classification loss $\mathcal{L}_{pcls}$.

3.3 Self-Supervised Outlier Synthesis

Ground truth object features and their corresponding pseudo-labels (Sec. 3.2) are used to obtain class-conditional Gaussian distributions that can be used for self-supervised outlier synthesis (SSOS), thus enabling decision boundaries between inliers and outliers. We implement the virtual outlier synthesis (VOS) technique by Du *et al.* [15], where the class-conditional Gaussians are formed from the penultimate layer features (after the shared head, Fig. 2).

A normal distribution for each object pseudo-class is constructed by having a mean $\boldsymbol{\mu}_k$ and a tied covariance $\boldsymbol{\Sigma}$ given by:

$$\boldsymbol{\mu}_k = \frac{1}{N_k} \sum_{i:\lambda_i=k}^{N_k} \mathbf{v}_{ij} \tag{4}$$

$$\boldsymbol{\Sigma} = \frac{1}{N} \sum_{k} \sum_{i:\lambda_i=k}^{N_k} (\mathbf{v}_{ij} - \boldsymbol{\mu}_k)(\mathbf{v}_{ij} - \boldsymbol{\mu}_k)^\top . \tag{5}$$

Subsequently, virtual outliers $\tilde{\mathbf{v}}$ are sampled from the normal distributions such that their probabilities are less than a value ϵ. Since ϵ is unknown, an approximation is carried out by sampling several features from $\mathcal{N}(\boldsymbol{\mu}_k, \boldsymbol{\Sigma})$ and taking the less likely sample as an outlier. To differentiate between normal and anomalous objects, the free energy is used as a confidence measurement, which is given by:

$$E(b_{ij}) = -\log \sum_{k=1}^{K} \exp(f_k w_k) \,, \tag{6}$$

where f_k are the pseudo-class logits of an object b_{ij} and w_k are learned weights assigning greater importance to some classes, following Du et al. [15]. A greater energy indicates a more anomalous object, whereas a lower energy score suggests an object conforming to the norm. From this energy, an MLP ϕ is used to predict an uncertainty score $\lambda_{ij} = \phi(E(b_{ij}))$, such that normal data has greater λ_{ij} values than outliers. During inference time, anomalies are detected by the predicted energy of objects detected by the OLN. With this strategy, SSOS regularises the feature representations of in-distribution objects to be compact, identifying anomalous objects by being far from all category clusters. Binary Cross Entropy is used for the classification of normal vs. anomaly, such that:

$$\mathcal{L}_{anomaly} = \frac{1}{N_n + N_o} \left(\sum_{i}^{N_n} \log\left(\phi(E(b_{ij}))\right) + \sum_{i}^{N_o} \log\left(1 - \phi(E(\tilde{\mathbf{v}}_i))\right) \right) , \tag{7}$$

where N_n is the number of normal data feature vectors and N_o is the number of outliers. The final loss function of our approach is thus given by:

$$\mathcal{L}_{OLN\text{-}SSOS} = \mathcal{L}_{RPN} + \mathcal{L}_{bbox} + \mathcal{L}_{mask} + \alpha\mathcal{L}_{pcls} + \beta\mathcal{L}_{anomaly} \,. \tag{8}$$

Feature Flow Synthesis (FFS) [28], a recent approach for outlier synthesis, uses a normalising flow function f that maps the complex in-distribution features into a simpler space for feature synthesis. We also explore using this technique in our method, and call this variant OLN-FFS. Specifically, $f : \mathbb{R}^d \to \mathbb{R}^d$ is a sequence of invertible bijective functions with parameters θ that transforms the object features $\mathbf{v}_i$ into a feature space $\boldsymbol{\xi}_i = f(\mathbf{v}_i)$ such that $p(\boldsymbol{\xi}_i) \sim \mathcal{N}(\mathbf{0}, I)$.

Virtual outliers $\tilde{\boldsymbol{\xi}}$ are sampled from this space and projected back to the object feature space via $\tilde{\mathbf{v}} = f^{-1}(\tilde{\boldsymbol{\xi}})$. During training, the log-likelihood of recovering $\mathbf{v}_i$ from f is maximised by adding the negative log-likelihood loss:

$$\mathcal{L}_{nll} = \frac{1}{N}\sum_{1}^{N} -\log(p_\theta(\mathbf{v}_i))\,, \tag{9}$$

where p_θ is the posterior likelihood and is given by $p_\theta(\mathbf{v}_i) = p(f(\mathbf{v}_i))|\det \mathrm{J}^{f,\mathbf{v}}|$, such that $\mathrm{J}^{f,\mathbf{v}_i}$ is the Jacobian matrix of f with respect to $\mathbf{v}_i$. This loss is added to Eq. (8) to form the OLN-FFS loss:

$$\mathcal{L}_{OLN\text{-}FFS} = \mathcal{L}_{OLN\text{-}SSOS} + \gamma\mathcal{L}_{nll}\,. \tag{10}$$

4 Experimental Setup

We evaluate OLN-SSOS and OLN-FFS on diverse datasets (Sec. 4.1) to show the effectiveness in detecting unseen anomaly objects. Sec. 4.2 reviews the performance metrics to evaluate our approach against the baseline methods and finally Sec. 4.3 gives an overview of our implementation details for reproducibility.

4.1 Datasets

In order to perform anomaly detection, datasets without object anomalies must be used for training. We evaluate diverse datasets from various image modalities (visible, infrared and X-ray). Following OoD works [13, 15, 28], we use the **PASCAL-VOC 2007 and 2012** [16] datasets with 20 object categories and the **Berkeley DeepDrive (BDD100K)** [68] dataset with 10 categories as ID while the OoD test sets consist on subsets of the MS-COCO [33] validation partition removing images containing in-distribution instances. To demonstrate the efficacy of our method in actual application scenarios, we also train our model on two X-ray security imagery datasets, **SIXray10** [39], a publicly available security inspection X-ray image with 5 object class labels and **Durham Baggage Full Image (DBF6)** [2] datasets containing 6 object class labels, and apply a leave-one-out contraband item anomaly detection formulation ('firearms' in SIXRay10, 'firearms' + 'firearm parts' in DBF6) to construct the in-distribution/OoD training and testing data partitions. Finally, we also use the publicly available **Long-term Thermal Drift (LTD)** [42] dataset, similarly applying a leave-one-out strategy (vehicle). A summary of the composition of these dataset formulations is presented in Supplementary Material.

4.2 Evaluation Metrics

OoD detectors focus on evaluating how accurately the predicted detections in the OoD dataset are flagged as outliers while keeping the in-distribution dataset detections with low false positive anomalies. However, this approach does not account for the anomalies recall. Motivated by this, we report class-agnostic MS-COCO [33] detection metrics to investigate the localisation performance. Since our method leverages an open-world detector, only the single-class average recall (AR) metrics are reported, specifically AR@10 (10 detections), AR@100 (100 detections), AR@S (small objects), AR@M (medium objects) and AR@L (large objects). Following convention, OoD detection performance is reported considering detections with an uncertainty score below a threshold such that 95% of in-distribution detected objects are above it (*i.e.*, they are flagged as normal). Following Du et al. [15], only in-distribution detections with a confidence score greater than an optimum threshold (that maximises the F1 score) are considered.

4.3 Implementation details

The OWOD sub-network of OLN-SSOS and OLN-FFS is implemented following the original OLN architecture [26], *i.e.*, a Faster-RCNN [46] (or Mask RCNN [21] for instance segmentation) detector with a ResNet-50 [22] backbone pre-trained on the ImageNet [27] and with no classification heads. In this sense, all ground truth class labels are ignored in order to account for learned pseudo-classes. UPC is carried out before each epoch using the ground truth object features extracted from the backbone (Sec. 3.2) using a RoIAlign layer with a 3×3 output size and 256 channels. These features are flattened, creating a $2,304$ vector representation of each bounding box. Pseudo-labels are obtained using the mini-batch k-means implementation of Sculley [55], using the resulting cluster centres as initialisation for the next epoch re-clustering. These pseudo-labels are used as ground truth classes to train the pseudo-label classifier (no background class is added), with a loss weight of 1 (α in Eq. (8)). The corresponding SSOS and FFS implementations follow the original settings, such that the anomaly classification module in Fig. 2 consist of a two-layer MLP with a ReLU activation and 512 hidden dimensions. We use a loss weight of 0.1 (β in Eq. (8)). OLN-FFS models use an *nll* loss weight of 1×10^{-4} (γ in Eq. (10)). For OLN-SSOS, outliers are chosen as the least confident out of $10,000$ class-conditional samples, while for OLN-FFS the samples are reduced to 300. We investigate the impact of the number of pseudo-labels, as well as the number of outlier samples in the ablation studies. Since DBF6 is the only dataset with available instance masks, we include variants using OLN-Mask [26]. We homogenise all implementations under the MMDetection [9] framework (VOS and FFS are implemented using Detectron2 [64]). The training regime is detailed in Supplementary Material.

Our proposed methods are compared against object-based OoD detectors SIREN [13], VOS [15] and FFS [28]. SIREN and VOS use a ResNet-50 backbone while FFS is trained with a RegNetX-50 [44]. For VOC and BDD in-distribution datasets, the original settings are used for the baselines. For DBF6, SIXRay10 and LTDImaging, all baseline methods are trained using similar configurations as in VOS, *i.e.*, minimum image size of 800 (except for LTDImaging, which is trained using an image size of 384), $10,000$ samples for virtual outlier synthesis and similar training recipe consisting of 18 epochs with a learning rate of 0.02 decaying by a factor of 10 in epochs 12 and 16. Following the original works,

Table 1: Anomaly detection metrics for PASCAL VOC and BDD datasets.

		In-distribution Test Set						COCO OoD Test Set					
	Method	**AR@1**	**AR@10**	**AR@100**	**AR@S**	**AR@M**	**AR@L**	**AR@1**	**AR@10**	**AR@100**	**AR@S**	**AR@M**	**AR@L**
PASCAL-VOC	SIREN [13]	23.9	56.5	59.7	37.3	52.5	68.6	9.0	19.2	19.9	4.5	8.8	35.2
	VOS [15]	23.8	56.3	59.5	37.0	52.2	68.6	**9.8**	**20.0**	**20.6**	4.1	9.7	**36.3**
	FFS [28]	**24.7**	**58.1**	60.9	36.2	54.3	**70.0**	**9.8**	19.2	19.6	**4.6**	9.9	33.8
	OLN-SSOS (Ours)	14.6	48.9	60.7	**44.3**	57.6	66.1	4.3	11.1	14.8	2.5	12.2	21.8
	OLN-FFS (Ours)	14.9	49.6	**61.3**	44.0	**58.4**	66.7	3.2	11.2	17.8	3.9	**19.2**	22.1
BDD	SIREN [13]	4.6	**32.3**	**51.8**	**37.6**	**63.2**	**85.6**	**3.0**	**9.2**	10.5	3.7	7.2	19.6
	VOS [15]	4.6	**32.3**	51.7	37.5	**63.2**	85.4	2.6	8.6	9.9	3.4	7.0	18.4
	FFS [28]	4.5	31.9	51.4	**37.6**	62.6	84.4	**3.0**	9.0	10.3	3.5	6.9	19.4
	OLN-SSOS (Ours)	**4.7**	27.9	45.9	29.6	59.7	83.4	0.5	1.6	3.5	3.0	6.0	1.4
	OLN-FFS (Ours)	4.6	27.0	44.1	27.8	57.8	81.7	1.6	6.0	**15.9**	**4.4**	**17.1**	**24.4**

outlier synthesis starts at epoch 12 for VOS and FFS. All experiments (our approach and the baselines) are trained using a single NVIDIA 2080Ti GPU.

5 Results

Tabs. 1 to 3 compare our method with state-of-the art OoD object detectors SIREN [13], VOS [15] and FFS [28]. While these methods are evaluated using the area under the receiver operating characteristic (AUROC) and the false positive rate at 95% in-distribution recall (FPR95), we instead report AR to assess detection performance (Sec. 4.2). Only in-distribution data is used during all training.

Tab. 1 presents the anomaly detection performance when training on the VOC and BDD100k datasets. Our method exhibits a great performance for the in-distribution PASCAL-VOC dataset despite being trained without class supervision (AR@100=61.3% vs. 60.9% for FFS) and a competitive OoD detection, with AR@100=17.8% compared with 20.6% for VOS. We highlight that our method does not use class labels, making it essentially unsupervised in this aspect, and further remark on the importance of such a class-agnostic focus for a technique where the prior class distribution is a strong prior that enables energy-based OoD detection [15]. Learning class distribution without class supervision is challenging given the intra-class variability (i.e., there are several modalities of the person class). Considering that VOC dataset is highly unbalanced, with several person instances, our method achieves a competitive performance. Similar in-distribution performance is observed on the BDD100k dataset. However, our approach significantly outperforms the baseline for OoD detection, with a maximum AR@100 of 15.9% (OLN-FFS) vs. 10.5% for the best baseline method (SIREN). This indicates that our approach achieves stronger OoD detection performance while maintaining moderately high accuracy in the original in-distribution object detection task. Furthermore, while the original implementation of FFS has greater FPR95 and AUROC metrics [28], this does not translate into localising anomalies, as evidenced by its low AR, especially for the BDD100k dataset that contains less classes than VOC. Our method overcomes this issue by sampling outliers from pseudo-class features with a normalising flow function, effectively detecting and localising anomalies, thus achieving stronger

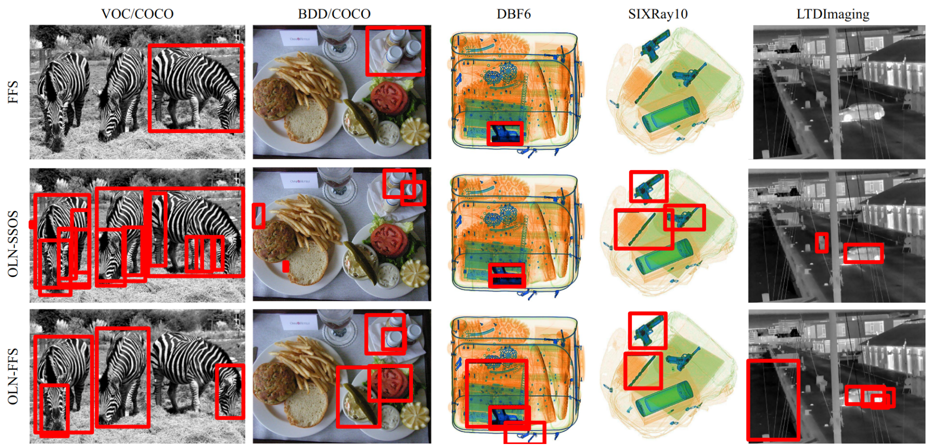

Fig. 3: The qualitative results exemplify the effectiveness of our proposed approaches (OLN-SSOS/OLN-FFS) in detecting OoD/anomalous objects (in red bounding boxes).

Fig. 4: Qualitative results obtained from the DBF6 dataset utilising OLN-SSOS Mask.

OoD detection performance. Qualitative results are presented in Fig. 3, with FFS chosen as the baseline since it obtains better metrics overall. It is observed that the baseline struggles to localise objects, while our method correctly localises and detects anomalies. Further qualitative results are included in the Supplementary Material.

Tab. 2 presents the results on the DBF6 dataset. We report detection performance in both bounding box and mask detection, demonstrating the extension of our approach capability to instance segmentation. Our method achieves significantly superior results on the OoD test set without affecting the in-distribution performance. We highlight that incorporating mask features effectively enhances OoD detection results (AR@10=55.1%, AR@100=58.9%). Qualitative results in Figs. 3 and 4 show that our method detects firearm and firearm parts as anomalies. Other electronics (tablets) are also detected, which are not present in the training set, underscoring the benefits of integrating mask features in SSOS.

Tab. 3 shows the results for SIXRay10 and LTDImaging. Here it is observed that while the baseline methods can perform in-distribution detection, they catastrophically fail to perform anomaly detection, with 0% AR. On the other hand, our approach offers in-distribution and OoD detection capabilities. For instance, OLN-SSOS achieves competitive in-distribution metrics, with

Table 2: Anomaly detection metrics for the DBF6 dataset.

	In-distribution Test Set						OoD Test Set					
Method	**AR@1**	**AR@10**	**AR@100**	**AR@S**	**AR@M**	**AR@L**	**AR@1**	**AR@10**	**AR@100**	**AR@S**	**AR@M**	**AR@L**
SIREN [13]	47.4	56.3	56.4	17.8	45.8	86.3	29.0	35.1	35.1	0.0	35.2	34.0
VOS [15]	43.9	54.3	54.4	17.6	44.4	82.7	23.5	32.8	32.8	0.0	31.7	49.1
FFS [28]	**49.0**	**56.5**	56.5	14.2	45.7	87.5	30.0	35.4	35.4	0.0	35.1	41.4
OLN-SSOS Box(Ours)	27.7	44.2	49.1	15.2	37.0	81.9	29.4	46.1	48.8	0.0	48.3	57.4
OLN-FFS Box (Ours)	29.2	45.8	51.5	16.6	40.1	83.1	15.6	35.9	46.3	0.0	46.0	51.4
OLN-SSOS Mask (Ours)	31.3	47.0	52.0	17.0	38.9	87.3	**34.1**	**55.1**	**58.9**	0.0	**59.0**	58.6
OLN-FFS Mask (Ours)	32.8	50.3	54.8	22.2	42.1	88.9	21.7	47.9	57.2	**10.0**	57.0	62.1

Table 3: Anomaly detection metrics for the SIXRay10 and LTDImaging datasets.

		In-distribution Test Set						OoD Test Set					
	Method	**AR@1**	**AR@10**	**AR@100**	**AR@S**	**AR@M**	**AR@L**	**AR@1**	**AR@10**	**AR@100**	**AR@S**	**AR@M**	**AR@L**
SIXRay10	SIREN [13]	47.8	63.3	63.7	10.0	62.8	64.7	0.8	0.8	0.8	0.0	0.0	0.9
	VOS [15]	48.2	63.6	63.6	0.0	62.2	65.0	0.0	0.1	0.1	0.0	0.0	0.2
	FFS [28]	**49.2**	**65.4**	**65.4**	60.0	**65.1**	**65.8**	0.8	0.8	0.8	0.0	0.0	1.0
	OLN-SSOS (Ours)	28.0	49.2	55.2	**70.0**	56.4	54.4	10.7	25.8	35.3	55.0	34.2	35.3
	OLN-FFS (Ours)	29.8	50.4	55.1	30.0	55.3	55.1	**12.1**	**27.3**	**35.6**	40.0	32.0	**36.0**
LTDImaging	SIREN [13]	5.9	34.2	**52.5**	**51.6**	75.1	-	0.0	0.0	0.0	0.0	0.0	0.0
	VOS [15]	**6.0**	**34.3**	**52.5**	**51.6**	75.1	-	0.0	0.0	0.0	0.0	0.0	0.0
	FFS [28]	**6.0**	34.2	**52.5**	**51.6**	**75.6**	-	0.0	0.0	0.0	0.0	0.0	0.0
	OLN-SSOS (Ours)	3.6	15.5	17.8	15.7	70.9	-	0.0	12.2	**18.2**	0.0	**17.6**	**62.9**
	OLN-FFS (Ours)	3.9	16.8	19.4	17.3	70.5	-	4.2	**12.3**	12.8	2.6	9.2	50.2

AR@100=55.2% vs. 65.4% of FFS, while having an OoD AR@100 of 35.6%. Similarly, while our methods have a significant drop in in-distribution performance on the LTDImaging dataset, they show great OoD detection, with an AR@100 up to 18.2%. Ablation studies (Sec. 5.1) show that the number of pseudo-labels may impact the performance, thus explaining the drop in in-distribution detection. Qualitative results in Fig. 3 show that the baseline methods cannot detect the objects since they are trained to only detect in-distribution objects.

Finally, the results in Tabs. 1 and 3 show superior OoD performance of OLN-FFS against OLN-SSOS for the VOC/COCO, BDD/COCO and SIXRay10 datasets, results on the DBF6 and LTDImaging show a drop in performance when using FFS. In this context, the work of [54] shows that invertible mapping, as in FFS, helps in forming high likelihood regions based on high-level object features. The OoD instances in DBF6 and LTDImaging are significantly different from the in-distribution classes such that these datasets might require stronger low-level features discrimination in order to achieve improved anomaly detection.

5.1 Ablation Studies

Fig. 5 shows the effect of varying the hyperparameters in OLN-SSOS. We present AR@10 and AR@100 metrics when varying the sampling size for SSOS and the number of pseudo-labels. Fig. 5a shows the effect of the outlier sampling size for the VOC/COCO dataset, ranging from 50 to $10,000$. A peak in performance is observed for 300 samples, identifying the boundary region between normal and abnormal samples. While relatively good performance is obtained for all other sampling sizes, it is crucial to identify the optimum choice for each dataset. Figs. 5b to 5d show the effect of changing the number of pseudo-labels for the DBF6, SIXRay10 and LTDImaging datasets. In general, the higher numbers of

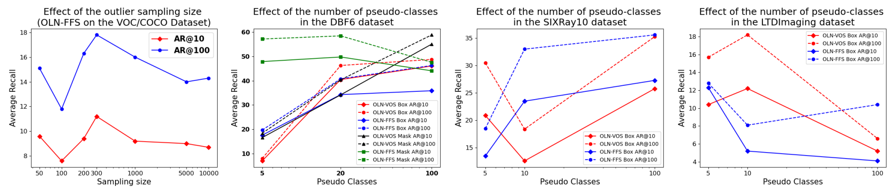

Fig. 5: Ablation Studies

pseudo-labels have a positive impact on performance (similar to DeepCluster [7]). This suggests the UPC may capture different modalities of data, indicating that over-segmentation helps in anomaly detection. In addition, this offers greater flexibility when designing OoD detectors since we are not tied by the number of ground truth classes. On the other hand, Fig. 5d shows a decrease in performance for a higher number of pseudo-classes on the LTDImaging dataset. Differently from the rest of the datasets, LTDImaging only contains three in-distribution classes (person, motorcycle and bicycle) with low inter-class variability, indicating that using several pseudo-classes might cause overfitting. Overall, the number of pseudo-classes must follow a detailed analysis of the dataset to achieve maximal performance. More ablations are available in the Supp. Material.

6 Conclusion

In this work, we introduce OLN-SSOS, an end-to-end open-world object-based anomaly detection, operating without class supervision to localise unseen anomalies within the training sets. Our method utilises an open-world object detector that learns object pseudo-labels, fitting object features into pseudo-class-conditional Gaussians to synthesise outliers from low-likelihood regions, enabling a better decision boundary between inliers and outliers during inference. We demonstrate the superiority of our approach over baseline methods, which solely rely on existing class-wise data for training in-distribution data.

Furthermore, we evaluate OLN-SSOS detection performance across different imaging modalities to assess its versatility. Our results reveal that while baseline approaches can detect in-distribution data, they struggle with anomaly detection, particularly in test sets where the anomalies are significantly different from the training classes. Conversely, our approach demonstrates its capability by extending to different imaging modalities (X-ray, infrared), showing improved performance in anomaly detection. This provides valuable insights into the generalisation capability of our proposed approach across varying imagery characteristics. Additionally, we extend our method to instance segmentation. The quantitative results illustrate the significant impact of using mask information, yielding better performance in detecting unseen anomalies while still maintaining moderately high in-distribution detection, showcasing the extension of our approach in instance segmentation.

Acknowledgement: Partially funded under Innovative Research Call in Explosives and Weapons Detection, 2023 (sponsored under the UK Government CONTEST strategy in partnership with US Dept. Homeland Security, Sci. and Tech. Directorate).

References

1. Akcay, S., Kundegorski, M., Willcocks, C., Breckon, T.: On using deep convolutional neural network architectures for automated object detection and classification within x-ray baggage security imagery. IEEE Transactions on Information Forensics and Security **13**(9), 2203–2215 (September 2018)
2. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: Asian conference on computer vision. pp. 622–637. Springer (2018)
3. Barker, J.W., Breckon, T.P.: Panda: Perceptually aware neural detection of anomalies. In: 2021 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2021)
4. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4183–4192 (2020)
5. Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C.: Improving unsupervised defect segmentation by applying structural similarity to autoencoders. arXiv preprint arXiv:1807.02011 (2018)
6. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: European conference on computer vision. pp. 213–229. Springer (2020)
7. Caron, M., Bojanowski, P., Joulin, A., Douze, M.: Deep clustering for unsupervised learning of visual features. In: Proceedings of the European conference on computer vision (ECCV). pp. 132–149 (2018)
8. Chan, R., Lis, K., Uhlemeyer, S., Blum, H., Honari, S., Siegwart, R., Fua, P., Salzmann, M., Rottmann, M.: Segmentmeifyoucan: A benchmark for anomaly segmentation. arXiv preprint arXiv:2104.14812 (2021)
9. Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X., Sun, S., Feng, W., Liu, Z., Xu, J., Zhang, Z., Cheng, D., Zhu, C., Cheng, T., Zhao, Q., Li, B., Lu, X., Zhu, R., Wu, Y., Dai, J., Wang, J., Shi, J., Ouyang, W., Loy, C.C., Lin, D.: MMDetection: Open mmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155 (2019)
10. Cohen, N., Hoshen, Y.: Sub-image anomaly detection with deep pyramid correspondences. arXiv preprint arXiv:2005.02357 (2020)
11. Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: International Conference on Pattern Recognition. pp. 475–489. Springer (2021)
12. Deng, H., Li, X.: Anomaly detection via reverse distillation from one-class embedding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9737–9746 (2022)
13. Du, X., Gozum, G., Ming, Y., Li, Y.: Siren: Shaping representations for detecting out-of-distribution objects. Advances in Neural Information Processing Systems **35**, 20434–20449 (2022)
14. Du, X., Wang, X., Gozum, G., Li, Y.: Unknown-aware object detection: Learning what you don't know from videos in the wild. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 13678–13688 (2022)

15. Du, X., Wang, Z., Cai, M., Li, Y.: Vos: Learning what you don't know by virtual outlier synthesis. arXiv preprint arXiv:2202.01197 (2022)
16. Everingham, M., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes (voc) challenge. International journal of computer vision **88**, 303–338 (2010)
17. Gaus, Y.F.A., Bhowmik, N., Isaac-Medina, B.K., Shum, H.P., Atapour-Abarghouei, A., Breckon, T.P.: Region-based appearance and flow characteristics for anomaly detection in infrared surveillance imagery. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2994–3004 (2023)
18. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)
19. Gudovskiy, D., Ishizaka, S., Kozuka, K.: Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 98–107 (2022)
20. Gupta, A., Narayan, S., Joseph, K., Khan, S., Khan, F.S., Shah, M.: Ow-detr: Open-world detection transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9235–9244 (2022)
21. He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask r-cnn. In: Proceedings of the IEEE international conference on computer vision. pp. 2961–2969 (2017)
22. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
23. Huang, Z., Huang, L., Gong, Y., Huang, C., Wang, X.: Mask scoring r-cnn. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 6409–6418 (2019)
24. Ionescu, R.T., Khan, F.S., Georgescu, M.I., Shao, L.: Object-centric auto-encoders and dummy anomalies for abnormal event detection in video. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7842–7851 (2019)
25. Joseph, K., Khan, S., Khan, F.S., Balasubramanian, V.N.: Towards open world object detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5830–5840 (2021)
26. Kim, D., Lin, T.Y., Angelova, A., Kweon, I.S., Kuo, W.: Learning open-world object proposals without learning to classify. IEEE Robotics and Automation Letters (RA-L) (2022)
27. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. Communications of the ACM **60**(6), 84–90 (2017)
28. Kumar, N., Šegvić, S., Eslami, A., Gumhold, S.: Normalizing flow based feature synthesis for outlier-aware object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5156–5165 (2023)
29. Lee, K., Lee, H., Lee, K., Shin, J.: Training confidence-calibrated classifiers for detecting out-of-distribution samples. arXiv preprint arXiv:1711.09325 (2017)
30. Lee, S., Lee, S., Song, B.C.: Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. IEEE Access **10**, 78446–78454 (2022)
31. Li, C., Han, Z., Ye, Q., Jiao, J.: Visual abnormal behavior detection based on trajectory sparse reconstruction analysis. Neurocomputing **119**, 94–100 (2013)

32. Li, W., Mahadevan, V., Vasconcelos, N.: Anomaly detection and localization in crowded scenes. IEEE transactions on pattern analysis and machine intelligence **36**(1), 18–32 (2013)
33. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: European conference on computer vision. pp. 740–755. Springer (2014)
34. Liu, W., Wang, X., Owens, J., Li, Y.: Energy-based out-of-distribution detection. Advances in neural information processing systems **33**, 21464–21475 (2020)
35. Liu, W., Luo, W., Lian, D., Gao, S.: Future frame prediction for anomaly detection–a new baseline. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6536–6545 (2018)
36. Liu, W., Li, R., Zheng, M., Karanam, S., Wu, Z., Bhanu, B., Radke, R.J., Camps, O.: Towards visually explaining variational autoencoders. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 8642–8651 (2020)
37. Lu, C., Shi, J., Jia, J.: Abnormal event detection at 150 fps in matlab. In: Proceedings of the IEEE international conference on computer vision. pp. 2720–2727 (2013)
38. Luo, W., Liu, W., Gao, S.: A revisit of sparse coding based anomaly detection in stacked rnn framework. In: Proceedings of the IEEE international conference on computer vision. pp. 341–349 (2017)
39. Miao, C., Xie, L., Wan, F., Su, C., Liu, H., Jiao, J., Ye, Q.: Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2119–2128 (2019)
40. Morais, R., Le, V., Tran, T., Saha, B., Mansour, M., Venkatesh, S.: Learning regularity in skeleton trajectories for anomaly detection in videos. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11996–12004 (2019)
41. Nguyen, T.N., Meunier, J.: Anomaly detection in video sequence with appearance-motion correspondence. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 1273–1283 (2019)
42. Nikolov, I.A., Philipsen, M.P., Liu, J., Dueholm, J.V., Johansen, A.S., Nasrollahi, K., Moeslund, T.B.: Seasons in drift: A long-term thermal imaging dataset for studying concept drift. In: Thirty-fifth Conference on Neural Information Processing Systems (2021)
43. Park, H., Noh, J., Ham, B.: Learning memory-guided normality for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14372–14381 (2020)
44. Radosavovic, I., Kosaraju, R.P., Girshick, R., He, K., Dollar, P.: Designing network design spaces. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (June 2020)
45. Ravanbakhsh, M., Nabi, M., Sangineto, E., Marcenaro, L., Regazzoni, C., Sebe, N.: Abnormal event detection in videos using generative adversarial nets. In: 2017 IEEE international conference on image processing (ICIP). pp. 1577–1581. IEEE (2017)
46. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems (2015)

47. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14318–14328 (2022)
48. Roy, P.R., Bilodeau, G.A.: Road user abnormal trajectory detection using a deep autoencoder. In: Advances in Visual Computing: 13th International Symposium, ISVC 2018, Las Vegas, NV, USA, November 19–21, 2018, Proceedings 13. pp. 748–757. Springer (2018)
49. Roy, P.R., Bilodeau, G.A.: Adversarially learned abnormal trajectory classifier. In: 2019 16th Conference on Computer and Robot Vision (CRV). pp. 65–72. IEEE (2019)
50. Rudolph, M., Wandt, B., Rosenhahn, B.: Same same but differnet: Semi-supervised defect detection with normalizing flows. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 1907–1916 (2021)
51. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: International conference on machine learning. pp. 4393–4402. PMLR (2018)
52. Sabokrou, M., Khalooei, M., Fathy, M., Adeli, E.: Adversarially learned one-class classifier for novelty detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3379–3388 (2018)
53. Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M.H., Rabiee, H.R.: Multiresolution knowledge distillation for anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 14902–14912 (2021)
54. Schirrmeister, R., Zhou, Y., Ball, T., Zhang, D.: Understanding anomaly detection with deep invertible networks through hierarchies of distributions and features. Advances in Neural Information Processing Systems **33**, 21038–21049 (2020)
55. Sculley, D.: Web-scale k-means clustering. In: Proceedings of the 19th International Conference on World Wide Web. p. 1177–1178. WWW '10, Association for Computing Machinery, New York, NY, USA (2010). https://doi.org/10.1145/1772690.1772862, https://doi.org/10.1145/1772690.1772862
56. Shi, Y., Yang, J., Qi, Z.: Unsupervised anomaly segmentation via deep feature reconstruction. Neurocomputing **424**, 9–22 (2021)
57. Sricharan, K., Srivastava, A.: Building robust classifiers through generation of confident out of distribution examples. arXiv preprint arXiv:1812.00239 (2018)
58. Sultani, W., Chen, C., Shah, M.: Real-world anomaly detection in surveillance videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6479–6488 (2018)
59. Tian, Z., Shen, C., Chen, H., He, T.: Fcos: Fully convolutional one-stage object detection. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 9627–9636 (2019)
60. Vernekar, S., Gaurav, A., Abdelzad, V., Denouden, T., Salay, R., Czarnecki, K.: Out-of-distribution detection in classifiers via generation. arXiv preprint arXiv:1910.04241 (2019)
61. Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 7464–7475 (2023)
62. Wang, H., Zhang, X., Yang, S., Zhang, W.: Video anomaly detection by the duality of normality-granted optical flow. arXiv preprint arXiv:2105.04302 (2021)

63. Wilson, S., Fischer, T., Dayoub, F., Miller, D., Sünderhauf, N.: Safe: Sensitivity-aware features for out-of-distribution object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 23565–23576 (2023)
64. Wu, Y., Kirillov, A., Massa, F., Lo, W.Y., Girshick, R.: Detectron2. https://github.com/facebookresearch/detectron2 (2019)
65. Wu, Z., Lu, Y., Chen, X., Wu, Z., Kang, L., Yu, J.: Uc-owod: Unknown-classified open world object detection. In: European Conference on Computer Vision. pp. 193–210. Springer (2022)
66. Yan, X., Zhang, H., Xu, X., Hu, X., Heng, P.A.: Learning semantic context from normal samples for unsupervised anomaly detection. In: Proceedings of the AAAI conference on artificial intelligence. vol. 35, pp. 3110–3118 (2021)
67. Yi, J., Yoon, S.: Patch svdd: Patch-level svdd for anomaly detection and segmentation. In: Proceedings of the Asian conference on computer vision (2020)
68. Yu, F., Chen, H., Wang, X., Xian, W., Chen, Y., Liu, F., Madhavan, V., Darrell, T.: Bdd100k: A diverse driving dataset for heterogeneous multitask learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 2636–2645 (2020)
69. Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L.: Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. arXiv preprint arXiv:2111.07677 (2021)
70. Zaheer, M.Z., Lee, J.h., Astrid, M., Lee, S.I.: Old is gold: Redefining the adversarially learned one-class classifier training paradigm. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14183–14193 (2020)
71. Zavrtanik, V., Kristan, M., Skočaj, D.: Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8330–8339 (2021)
72. Zhao, X., Ma, Y., Wang, D., Shen, Y., Qiao, Y., Liu, X.: Revisiting open world object detection. IEEE Transactions on Circuits and Systems for Video Technology (2023)
73. Zohar, O., Wang, K.C., Yeung, S.: Prob: Probabilistic objectness for open world object detection. In: CVPR (2023)