2021 IEEE Fourth International Conference on Artificial Intelligence and Knowledge Engineering (AIKE) | 978-1-6654-3736-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/AIKE52691.2021.00013

A Survey on Open Set Recognition

1st Atefeh Mahdavi
Dept. of Engineering and Sciences
Florida Institute of Technology
Melbourne, USA
amahdavi@fit.edu

2nd Marco Carvalho
Dept. of Engineering and Sciences
Florida Institute of Technology
Melbourne, USA
mcarvalho@cs.fit.edu

***Abstract*—Open Set Recognition (OSR) is about dealing with unknown situations that were not learned by the models during training. In this paper, we provide a survey of existing works about OSR and distinguish their respective advantages and disadvantages to help out new researchers interested in the subject. The categorization of OSR models is provided along with an extensive summary of recent progress. Additionally, the relationships between OSR and its related tasks including multi-class classification and novelty detection are analyzed. It is concluded that OSR can appropriately deal with unknown instances in the real-world where capturing all possible classes in the training data is not practical. Lastly, some new directions for future research topics are suggested.**

***Index Terms*—machine learning, classification, open set recognition, multi-task learning, risk of the unknown**

I. Introduction

In OSR, only a limited number of known classes are available at the time of training the model and the possibility of unknown classes never seen at training time emerges in the test environment. In such a setting, the unknown classes and their risk should be considered in the algorithm. Such systems require not only to identify and discriminate instances that belong to the source domain (i.e., the seen known classes contained in the training dataset) but also to reject unknown classes in the target domain (classes used in the testing phase). Until recently, the success of almost all machine-learning-based systems has been obtained by conducting them on "closed-set"classification tasks. In such systems, the source and target domains are assumed to contain the same object classes and the system is only tested on known classes that have been seen during training. Different from the "closed set"setting, a more realistic scenario is solving real-world problems consisting of an "open set"of objects. With the advent of building intelligent systems and utilizing machine-learning-based systems, a wide range of applications require robust AI methods. Handling the "unknown unknowns"can be considered as one of the approaches that enable the system to act robustly in the face of limitations and unmodeled aspects of the world. Ignoring unknown objects causes improper development of the systems and limits their usability. However, building a correct and complete model for the recognition/classification task in the real dynamic world poses multiple challenges as anticipating and training all possible examples of unknown objects are prohibitive and the model may fail when assessed in testbeds.

Emerging real-world recognition systems require OSR to recognize unknown inputs and learn them when needed. OSR is a challenging task in a large number of safety environments where even a small fraction of errors on unknowns could place human lives at risk, such as a self-driving car defect or robotic surgical assistants with flaws in perception and execution. Additionally, large-scale recommendation systems deployed in social environments, where content and user preferences change dynamically call for adaptive techniques that learn fast, in an online fashion [1, 2, 3]. Moreover, real-world robots can expand their knowledge if they will be able to detect unknown objects, discover the need to learn about them and learn them continuously. An automatic face recognition system that is usually encountered with unknown individuals is another domain to be deployed in open-universe scenarios. Another area where OSR can be a solution is malware classification for cyber-security. This domain is faced with the challenge of incomplete knowledge of the training data because of emerging novel types of malware. The ever-changing nature of malware, as the intruders are continuously altering network attacks to bypass the existing detection solutions, calls for the development of autonomous countermeasures and the recognition of novel malware classes.

This paper is organized as follows: First, we briefly differentiate OSR from multi-class classification and novelty detection problems and discuss their limitations. Then, we propose two broad categories for OSR algorithms and details of the important studies under each category. The categories described here encompass statistical-based and deep-neural-network-based algorithms. Table 1 represents the classification of these categories and their sub-categories formed by the existing methods. In the last section, we provide an overall conclusion for this review.

II. OSR vs. Multi-class Classification and Anomaly Detection

OSR is referred to as a classification-based task. Most of the approaches to OSR were formed based on regular classifiers due to their closeness to the classification task; however, the adaptation of a classifier which is valid for OSR is not always possible. Classification and anomaly detection are the closest relatives to OSR. The relationship between OSR and these related areas is summarized in Table 2.

978-1-6654-3736-3/21/$31.00 ©2021 IEEE
DOI 10.1109/AIKE52691.2021.00013

TABLE I
A GLOBAL PICTURE OF THE EXISTING OSR METHODS.

Statistical Models				Deep Neural Networks		
Rejection-adapted SVM	**Sparse Representation**	**Distance-based**	**Margin Distribution**	**Adversarial Learning**	**Background-class-based Modeling**	**Others**
[4] [5] [6] [7] [8] [9] [27] [28] [29]	[10]	[11] [12] [13] [14]	[15] [16]	[17] [18] [19] [30][31] [32]	[20] [21] [22] [23]	[24] [25] [26] [33] [34][35] [36][37]

Multi-class Classification: In a conventional multi-class classifier, because of the closed set assumption, all inputs are labeled and classified into one of the known classes observed during training. More precisely, in the closed set classification task, the learner only has access to a fixed set of known classes $C = \{L_1, L_2, ...L_M\}$ and constructs an M-class classifier during the training phase. The resulting classifier is tested on the data from only the M classes. However, a problem emerges with the appearance of a test sample from an unknown class which does not belong to any of the known classes. Thus, the most likely class for an input observation is always provided and an unknown will wrongly be recognized as a sample belonging to one of those pre-defined classes. In OSR; however, knowledge of the entire set of possible classes cannot be considered during training. The classifier is allowed to predict classes from the set of $C^{'} = \{L_1, L_2, ..., L_M, L_{M+1}, ..., L_{M+\Omega}\}$, where classes L_{M+1} through $L_{M+\Omega}$ cover all unknown classes not observed during training but which appeared at query time. A test sample may be predicted to belong either to one of the known classes $c_i \in C$ or to an unknown one.

The difference between OSR and traditional classification is visualized in Fig. 1. The decision boundaries in Fig. 1.a are considered by training a traditional Nearest Class Mean (NCM) classifier on three different known classes illustrated by diamonds, circles, and squares and the unknown inputs represented by stars. Fig. 1.b demonstrates the distribution of original dataset in the open space when zooming out from the closed three-class model. Having incomplete knowledge of the entire set of possible classes, this classifier assigns class labels from the closed training set to an unlimited region. Therefore, at the classification time, the unknown inputs in the open space will be misclassified. On the other hand, OSR discriminates known samples and limits the scope of decisions by the support of the training data (see Fig. 1.c).

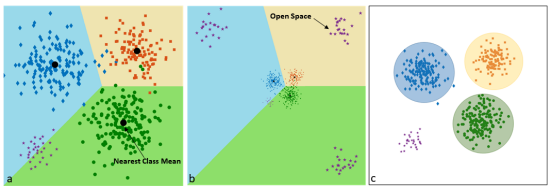

Fig. 1. A global picture of the existing OSR methods

Classification with Reject Option: There are many approaches regarding classification with a reject option in the literature which have been adjusted to support open sets. The greatest part of rejection-adapted approaches rested upon variants of Support Vector Machine (SVM) classifiers with the ability to reject observations, and the one-class classifiers based on support vectors. Although such techniques are related to OSR in the sense of rejecting an input, they have different reasons to do rejection actions. Classifiers with rejection options focus on the ambiguity between classes to reject an uncertain input of one class as a member of another one and minimize the distribution mismatch between the training and testing domains, while OSR rejects an input because of not belonging to any of the known classes. Therefore, rejecting uncertain inputs in such classifiers protects misclassification but is not enough to handle unknowns. They have infinite positively labeled open space and infinitely open space risk and thus are not able to solve OSR problems formally. In these techniques, unknowns often appear to be uncertain and are labeled with confidence. In contrast, OSR supports rejecting the unknown object by discovering the acceptable amount of uncertainty and searching among any of the known classes to identify if the true class exists. In other words, the set of possible outcomes of predictions is an important difference between OSR and a typical multi-class classifier. For example, in margin classifiers like SVMs, confidence is evaluated in terms of an associated distance to the decision boundary given for each example. The goal of SVMs is to find an optimal hyperplane to classify and separate the classes of training samples. The hyperplane defines half-spaces and divides examples of the separate categories by maximizing the distance between itself and the nearest training points. In such classifiers uncertainty is high near the decision boundary and confidence will be increased with distance from the decision boundary; the farther an input is from the margin, the more confident one can be that it belongs to the known classes. Thus, an unknown far from the boundary is incorrectly labeled and will be incorrectly classified with very strong evidence. For example, in Fig. 2, a plane found by the SVM separates bicycles and airplanes and maximizes the SVM margin making "airplane" a half-space. An unknown ("?") far from the training data will be misclassified and likely be labeled "airplane".

For such classifiers that use observation-to-margin distance as the only information to identify unknowns, resting on a threshold is not enough for discovering the hidden unknown classes. Moreover, due to incomplete information about unknown classes, selection of the decision threshold depends merely on the knowledge of known classes, and the decision score calibration is processed implicitly by closed set assumptions. Therefore, OSR cannot use rejection-adapted SVM as a good option, although it outperforms a multi-class SVM which

TABLE II
RELATIONSHIP BETWEEN OSR AND THE RELATED AREAS.

Settings	Training Data	Testing Data	Tasks
Traditional Classification	Known	Known	Classifying known data
Anomaly/Outlier Detection	Known	Known/Unknown	Identifying rare items
Open Set Recognition	Known	Known/Unknown	Classifying known data and rejecting unknowns

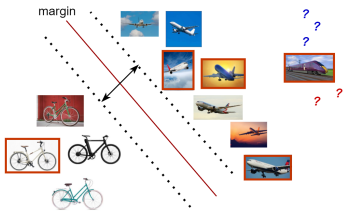

Fig. 2. Two-class SVM classifier. Images with an orange box are from testing and the rest of images are from training. Testing images can be known ("bicycle", "airplane") or unknown ("train," "?").

strictly assigns a label to the known.

Anomaly Detection: On the other hand, some people argue that identifying novel classes and detecting anomalies sometimes can solve the OSR problem. Although these methods referred to the problem of identifying unknown data and have been a good start for OSR, the problem setting is different from that of OSR. These techniques alone are not sufficient for creating a balance between the risks of the unknown and multiclass recognition for OSR, leading to poor performance.

III. A CATEGORIZATION OF OSR TECHNIQUES

A. Statistical approaches

[4] introduced the first formalization of OSR by balancing open space risk R_O associated with labeling data that is far from known training samples against minimizing empirical risk R_ϵ over training data. By assuming f to be a measurable recognition function, open space risk $R_O(f)$ for known class k is described as the following:

$$R_O(f) = \frac{\int_O f_k(x)dx}{\int_{S_k} f_k(x)dx} \tag{1}$$

This formalization provides the proportional value of positively labeled open space O against to the total measure of S_k consisting all of the known positive training samples $x \in k$ as well as the positively labeled open space. This paper argued that the essential element of OSR is finding the recognition function f, where $f(x) > 0$ indicates the positive recognition of the class k of interest. This function is defined as a minimization of the open space risk to capture the risk of labeling the unknown samples as known, beyond the sensible recognition of the training data, as follows.

$$argmin_{f \in H} \{R_O(f) + \lambda_r R_\epsilon(f)\} \tag{2}$$

where λ_r is the regularization tradeoff between open space risk and empirical risk. This study proposed a "1-vs-set Machine" which consists of two parallel hyperplanes. The proposed formulation with a linear kernel balances empirical and open space risk by exploiting a second hyperplane from the marginal distances of a 1-class or binary SVM. The main hyperplane is a base SVM which defines half-spaces and aims at maximizing the margin. The second hyperplane is added in such a way as to minimize the positive labeled region bounded between two planes and handle open space risk. This method defines a definition that generally describes region of known classes for each individual binary SVM, however, it lacks the procedure of distance measurements. It even does not clarify the space for measurements of such distances. Therefore, it cannot bound the space that each known class belongs to that leads to the existence of open space risk.

There have been many more attempts over the past years to address open space risk for training OSR models. Followup works [5, 6] were inspired by the fact that leveraging Extreme Value Theory (EVT) on the SVM decision scores provides better performance than exactly applying the raw score values. Both approaches have proposed EVT-based SVM calibration techniques to enable the SVM-based classification to deal with an open-set setting. Other OSR algorithms such as [15, 24, 25] also include EVT to analyze the association of a data point with an unknown class. Extreme value modeling has been increasingly used to analyze post-processing scores and enhance the performance of OSR. This theory is meant to study a level of confidence by determining the fraction of objects deviating from the expected value. As extreme values appear in the tails of the distributions, EVT examines the distribution tails and aims to predict the probability that a given sample is an extreme value. For OSR, EVT models the probability distributions of the match and non-match recognition scores and the rejection threshold is usually estimated from the overlap region of extreme values found in the tails of probability distributions. Unlike SVM, which divides all the space with hyperplanes to define sections and allocate them to one of the current classes, [6] applied EVT and Weibull distributions to build such hyperplanes without dividing the whole space. This work introduced the idea of a compact abating probability (CAP) model based on a one-class classifier which, when thresholded, can further limit the open space risk. The labeled region is limited and the open space risk is minimized if the value of the probability of class membership is decreasing in all directions as samples pull out of the training data and move towards

the open space. Distribution of decision scores for unknown recognition is considered by extending "1-vs-Set Machine" to W-SVM (a Weibull-calibrated non-linear classifier). This algorithm yields better modeling for a binary SVM at the decision boundaries.

Based on this intuition, [5] introduced a variant of W-SVM that is called Support Vector Machines with Probability of Inclusion (PISVM). This algorithm formulates the multi-class OSR problem as one of the modeling positive training data at the decision boundary. An SVM with RBF kernel is utilized as a binary classifier for each class and trained by the One-vs-All approach, where the samples of the remaining classes are assumed as negative. It models the unnormalized posterior probability of inclusion for multiple classes as a basis to reject unknown samples. Then, it fits probability distributions consistent with the statistical EVT, leveraged on the decision scores from the positive training samples. For a given sample, a class is chosen whose decision value makes the maximum probability of induction. The sample is recognized as unknown if that maximum is under a predefined threshold. Although the proposed algorithm is more accurate than W-SVM, it did not always confine open space risk, the issue that occurred with a regular SVM.

In spite of being a recent research focus, EVT for OSR does not merely guarantee a bounded open space, as PISVM [5] does not always bound open space and W-SVM [6] relies on one-class models to bound open space rather than counting on EVT models. Compared to EVT-based models, a simpler algorithm called Specialized SVM was proposed by [27] recently. This algorithm bounds the represented space for known categories and provides a finite risk of the unknown if using an RBF kernel and limiting the bias term to be negative. Additionally, the proposed EVT-based calibration of 1-vs-rest RBF SVMs modeling in both W-SVM and PI-SVM has two deficiencies. The first deficiency is that it is not ideal for OSR which requires incremental updates. Supporting incremental learning is a principal goal in designing an algorithm for OSR, especially one that is constantly used over a long period of time. This approach cannot add novel detected objects and tune the model to enhance the fit as a new item arrives. So it is not able to learn the model incrementally. The second deficiency is that it does not address the fundamental issue of choosing thresholds, which requires prior knowledge in such threshold-based classification models. However, the authors of the last two papers recommended choosing the thresholds according to the problem openness, which is not reasonable since the openness is not usually known in the corresponding problem.

To tackle these deficiencies [28] formulated the probabilistic open space SVM (POS-SVM) which rests on a one-vs-all binary SVM. An individual reject threshold for each of the known classes is computed and optimized by a validation set. Another possible approach more appropriate for the open-world with incremental learning capabilities proposed a Nearest Non-Outlier (NNO) algorithm [11]. NNO adapts the Nearest Class Mean Classifier (NCM), the basis of most open-set classifiers, for OSR by using non-negative combinations of abating distance. This work is built on the concept of a CAP model; however, it generalizes the model to gain zero open space risk by applying a threshold on any non-negative combination of abating functions. NCM represents the classes by the mean feature vector of their components, and a test sample is set to a class with the closest mean using Euclidean distance between the class mean and the test feature vectors. The NNO algorithm was inaccurate because of using thresholded distances from the nearest class mean. Additionally, it does not tune the rejection threshold automatically as new classes arrive and the problem evolves. So this algorithm does not properly model the dynamic nature of open-world recognition. To mitigate this problem, [12] used the Hoeffding bound to incrementally update the threshold for an unknown class, instead of estimating it from an initial set of known classes and keeping it fixed as previously used in [11]. [14] represented Nearest Centroid Class (NCC) which is incremental learning and built upon the NCM algorithm. Instead of using the class mean for each class member, this model is based on a series of closest neighbors of the centroid class. In spite of its similarity with NNO in terms of using multiple centroids, the proposed model addresses the issue of the new classes being added incrementally related to NNO and updates information for a class ball. During training, this algorithm attempts to create the boundary region for each known class. Each class is a set of balls centered at class centroids where each ball represents a number of its data points. An observation is treated as an unknown when not any of the nearest class boundaries support it.

Extreme Value Machine (EVM) [15] as a probabilistic framework for open set classification also considers Weibull distributional information when learning recognition functions. EVM is the first classifier to perform a nonlinear RBF approach motivated by EVT and provides a more powerful representation model for OpenMax which will be discussed in section III-B. Using CAP models, EVM is able to bound open space. However, this approach has drawbacks with regard to the choice of the threshold which controls the open set classification error and more important, strongly relies on the relative arrangement of the known classes. EVM assumes that the behavior of the unknowns can be inferred by the geometry of the known classes, and thus the recognition task may fail when the known and unknown geometries of classes are different. To overcome these limitations, two robust algorithms [16] derived from EVT that do not rely on the geometry of the observed data. These classifiers, called generalized Pareto distribution (GPD) and generalized extreme value (GEV), utilize the intuition that new points to be classified as known or unknown are more likely to be unknown if they are far away from the training data. Moreover, these algorithms are efficient to update upon arising new training data.

[13] proposed the Nearest Neighbor Distance Ratio classifier, which in turn, is a multiclass open-set extension for the Nearest Neighbor (NN) algorithm and is referred to as Open Set NN (OSNN). The OSNN first finds the nearest and second nearest neighbors y and z regarding a test sample t in order

that $\omega(y) \neq \omega(z)$, where $\omega(s) \in L = \{l_1, l_2, ..., l_n\}$ represents the class of sample s and L is a set of training labels. Then, this classifier calculates the similarity scores' ratio and applies a threshold to recognize sample t as unknown having low similarity. This ratio is defined by $r = d(t, y)/d(t, z)$, where Euclidean distance of two samples s and $s^{'}$ is shown by $d(s, s^{'})$. Additional works like Assign-and-Transform-Iteratively (ATI) [29], LACU (Learning with Augmented Class with Unlabeled data) framework [7] and Separate to Adapt (STA) [8] require the help of unknown source samples. In an experimental study on the open-set classification models for web genre identification (WGI) setup, [9] examined one-class SVMs and Random Feature Subspacing Ensembles (RFSE) models. With respect to this fact that most of the complementary information to differentiate known from unknown samples is placed in the tail of a distribution, modeling the tail of match and non-match error distributions can help to find the optimal threshold for a given recognition model. Inspired by this intuition, [10] extended the Sparse Representation-based Classification (SRC) algorithm to OSR. This algorithm models the tails of these two residual errors using EVT. The identity of an unknown test sample and open-set identification is determined by getting the confidence score for that sample and hypothesis testing.

B. Deep neural network-based algorithms

Following the extensions of traditional classification algorithms for OSR, there is a considerable amount of research in developing Deep Neural Networks (DNNs) for OSR in the literature. However, with the shift to deep networks, which combines learning features and learning the classifier, the performance of the system for OSR is still far from optimal.

Researches have addressed this problem by thresholding on the Softmax scores. the Softmax function represents the probability that the sample x is labeled with class n. Due to its closed nature, a deep network links an unknown sample to the class with the maximum score given by Softmax, leading to misclassification of that sample. It seems that for an unknown sample this function produces low probability for all the classes so that thresholding on the output probability can help to reject the unknowns. However, combining a deep network with thresholded probabilities determines uncertain predictions which are a small part of unknown inputs. As a consequence, thresholding Softmax is not enough to detect fooling or adversarial examples. Fooling examples target the desired class and seek to increase the corresponding probability of that class. These artificially constructed examples are fully imperceptible to humans, but the classifier sees them as a member of the desired classes (Fig. 3.c from [24]). A more restrictive case is rejecting an adversarial example– a visually similar input to the training dataset with small but intentional perturbations, such that it is mislabeled by a classifier as an entirely different class with high confidence (Fig. 3.d). The difficulty level of rejecting adversarial examples depends on how close the example is to the target class. If an adversarial example is produced from a nearby class, it will fail to be rejected as an unknown. However, if this example is generated from a faraway target class, it will be rejected as an unknown due to a remarkable difference in the output scores. That is why most of the studies proposed for OSR do not consider these examples in their experiments.

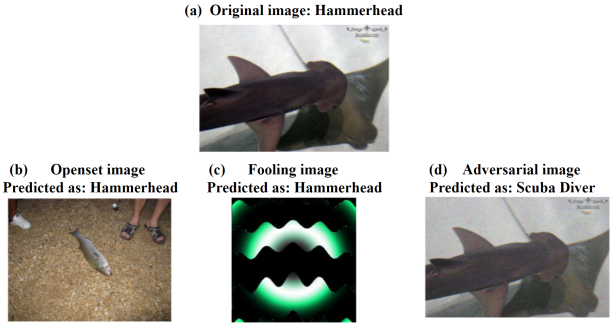

Fig. 3. Examples of an original, open set, fooling, and adversarial images. (b) A real image from an unknown category which is mapped to the class with the maximum response provided by Softmax. (c) A fooling input image which is unrecognizable to humans, but DNNs believe with high certainty to be a Hammerhead. (d) An adversarial image specifically constructed from hammerhead to scuba to fool DNNs into making an incorrect detection.

Moreover, a more effective rejection solution than thresholding softmax is using a garbage or background class which has dominated most of the modern detection approaches like [21, 22, 23]. Such background-class-based modeling can tackle the problem of unknowns in neural networks by adding another class as representative of unknown samples during training. However, this approach has limitations in the real world with infinite negative space of infinitely many unknown inputs to be rejected. Recently, [20] combined SoftMax with the Entropic Open-Set and Objectosphere losses considering the background and unknown training samples. These losses increase SoftMax entropy for unknown inputs while minimizing the Euclidean length of deep representations of unknown samples. This modification increases separation in deep feature space and improves the handling of background and unknown classes.

The OpenMax proposed by [24] was the first deep open-set classifier without using background samples. Since then, few deep open-set classifiers have been reported. OpenMax does not directly focus on the recognition of adversarial inputs, although it supports the rejection of fooling and unknown images. Although Openmax is more robust than Softmax to adversarial examples and outperforms networks with thresholding SoftMax, it does not provide robustness to sophisticated adversarial construction techniques. OpenMax uses the EVT model built from the positive training samples to define a per-class CAP model that can bound the risk of open space to reject unknown inputs by thresholding the model. First, Mean Activation Vector (MAV) is computed for each class separately over only correctly classified training examples. Then a Weibull distribution will be fitted to each class on the largest distances between the MAV class and positive training instances (Fig. 4).

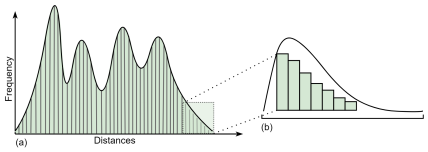

Fig. 4. (a) Class distance distribution based on measured distance between each correctly classified training example and the associated MAV. (b) A Weibull distribution related to a certain number of the largest such distances.

OpenMax does not enhance the feature representation for better unknown detection. The challenge of using the distance from the MAV is that the class instances are not projected directly around the MAV by using normal loss functions, such as cross entropy. Moreover, because the testing distance function is not used during training, it might not necessarily be the right distance function for that space.

Recently, there exist also abundant research on OSR based on the scheme of Generative Adversarial Networks (GANs). A GAN which recently stands out among various deep neural networks consists of a generator and a discriminator. Generally, the generator produces synthetic samples and the discriminator learns to decide if a sample is obtained from the generator or the real dataset. G-OpenMax [17] extends OpenMax in adversarial settings and applies GANs to generate unknown instances. These synthetic instances are utilized as an extra training label apart from known labels to adjust the classifier and estimate the probability of unknown classes. Along with a similar motivation, another GAN-based approach which is more effective than G-OpenMax for OSR was proposed by [18]. This strategy, which is called counterfactual image generation, searches for synthetic images by adopting an encoder-decoder GAN technique. These images, referred to as counterfactual-images, are a member of unknown classes but they look like known classes. Using the GAN framework, another work [19] aims at generating synthesis data which were served as fake unknown classes for the classifier to make it robust against real unknown classes. [30] proposed the adversarial sample generation (ASG) framework that produces unseen class data. Inspired by GANs, [31] proposed a novel model called Open-GAN. In this model, fake target samples are constructed from the generator automatically. Afterwards, the discriminator is modified to adapt multiple classes together with an unknown class. Another deep learning method [32], Open Set Back Propagation (OSBP), utilizes adversarial training for a more challenging open-set framework which does not require unknown source samples. This approach trains a feature generator to extract features that distinguish known target samples from unknown. A DOC (deep open classifier) proposed by [34] was compared with OpenMax and represents better performance. One issue related to OpenMax is classifying samples which are difficult to handle as these samples are mostly classified as members of unknown classes. The proposed classifier addresses this issue where the SoftMax layer is replaced by a one-vs-rest final layer of sigmoid activations. The network is trained using a novel loss function to perform joint classification and unknown detection and reduce the open space risk. They show that the risk of open space is reduced further for rejection and the algorithm is improved further by tightening the sigmoid functions' decision boundaries with Gaussian fitting. One possible drawback of this approach would be the lack of compact abating property of the sigmoids which may cause the problem of unbounded open space risk when they are activated by an infinitely distant input from all of the training data.

Another methodology for OSR is using a weightless neural network, denominated WiSARD. Compared to various classifiers, WiSARD does not rely on prior knowledge regarding data distribution, which is usually unavailable in OSR tasks. The proposed model assigns fitness scores to each class and evaluates how well a given observation matches the previously stored knowledge. This classifier applies such a fitting level for rejection according to the similarity rating and proximity between corresponding features. Computing score thresholds, this paper [26] developed a rejection-capable WiSARD to identify whether observations pertaining to the class with the highest score or the best score is below the defined threshold, and then it is considered as an outlier. Following that, after proposing some exploratory results, a fully developed methodology is detailed in [33]. This paper investigates how to adapt the WiSARD classifier for OSR by carrying out detailed distance-like calculations and defining the rejection thresholds at the training.

Until recently, almost all existing deep open-set techniques included standard neural networks which are trained in a closed set environment and different activations which are analyzed to infer unknowns. However, relying on discriminative features of known classes in such systems causes specialization of learned representations to known classes and is not useful to represent unknowns. In contrast, some approaches enhance the learned representation to keep useful information to jointly perform known classification and unknown detection. Classification-Reconstruction learning for Open-Set Recognition (CROSR) is the novel framework proposed by [25]. This is the first neural network architecture which involved hierarchical reconstruction blocks and trained networks for joint classification and reconstruction of input samples. The proposed system consists of a closed-set classifier which exploits learned prediction y for known class classification, and an unknown detector which uses a reconstructive latent representation z together with y for unknown detection. This study which considers deep representation learning is similar to [10] in terms of sharing the idea of reconstruction-based representation learning; however, [10] uses a single layer linear representation. The work in [36] combined a shared feature extractor that provides a latent space representation of an input image, along with a decoder and a classifier. Reconstruction errors from the decoder network are utilized to reject samples from the unknown classes. Another work [35] proposed an algorithm using class conditional auto-encoders. In this method, the training procedure is divided into two parts to learn open-set identification scores. The first part, closed-

set classification, is learned by an encoder using the traditional classification loss and the closed-set training setting, while a decoder reconstructs conditioned on class identity to train an open-set identification model and accomplish the second part of the training. In [37], the known and unknown classes are first distinguished based on entropy measurement and training the model on a modified cross entropy loss by dedicating a low and high cross-entropy for known and unknown classes, respectively. Then it uses the weighted square difference loss to assign unlabeled target samples to known classes based on the likeliness.

IV. Conclusion and Discussion

OSR arises due to an increased demand for a good classification or detection system, and thus a survey on this topic is particularly important. This survey tries to provide a structured and comprehensive overview of contemporary research on OSR. By comparing existing OSR techniques under two broad categories and discovering their limitations, we hope that this paper facilitates the promising subsequent research and a better understanding of this topic.

Based on the literature review, some related topics for future research can be discussed as follows. As mentioned previously, the classification setting is usually a classic closed-set problem. In this setting, we face the risk of open space and misclassification of an unknown sample falling into over-occupied space divided for the known classes. An appropriate understanding of the nature and the underlying structure of the data can help us to arrange the known classes in a more compact form and limit the open space. The clustering technique is to create meaningful groups of the given samples based on the similarity that can improve the exploration of the data information and the generalization ability of classification learning. Thus, designing a learning framework that combines clustering and classification tasks can overcome the problem of over-occupied space. However, all current existing simultaneous learning clustering and classification algorithms are designed for closed-set problems. Therefore, designing such a framework under open-set assumptions could be a promising direction. In OSR, we do not know what all the classes are, so modeling unknown classes is not possible. One way to enhance the learning ability and robustness of the classifiers is by adopting the idea of adversarial learning. This novel technology is to generate examples that are close to the training data set but different from any training category. The fake data can be considered as unknown samples and used for the regularisation of the classifiers. The key factors that are worth further exploring are: how to generate valid examples, how to omit artificial selection by automatic construction of synthesis samples in the training process, what level of similarity should be designed between the distribution of synthesized examples and known samples. Although less similarity results in easier discrimination between known and synthesized samples, the performance of the classifier for unknown classes might be also lower. On the other hand, a high level of resemblance leads to achieving better discriminability during the training of the classifier. Most of the OSR methods use threshold-based strategies in which the threshold is selected using the knowledge of the known classes. Thus, having no prior knowledge about unknown classes and defining a constant global threshold that no longer changes during the testing time, lead to OSR risk. That is, it misclassifies the unknown samples when they fall into the specified space for the known classes. Therefore, there is room to effectively determine a more robust way to select the threshold. For example, modifying the threshold based on the information of unknown classes received at the test time can improve the robustness of OSR techniques. Another investigation would be the robust selection of the tail's size while applying EVT to model the data distribution's tail. This strategy can be helpful at the critical problems where outliers in known classes and unknown samples appear simultaneously in the testing phase. Moreover, adversarial images are not necessarily isolated regions in the input space. They tend to stay adversarial across classifiers; consequently, they appear near a training sample and occupy contiguous areas in the input space. This problem is different from standard open space risk and presents a more difficult challenge towards existing OSR techniques, as they have not considered adversarial images in their experiments. It is, therefore, of great significance to design a framework for OSR that also takes the detection of adversarial images into account that can bring numerous benefits for the robustness and security of deep neural networks.

References

[1] K. Mahadik, Q. Wu, S. Li, and A. Sabne, "Fast distributed bandits for online recommendation systems," in *Proceedings of the 34th ACM international conference on supercomputing*, 2020, pp. 1–13.

[2] S. Li, F. Hao, and H.-C. Kim, "Online social media review mining for living items with probabilistic approach: A case study," *Smart Media Journal*, vol. 2, no. 2, pp. 20–27, 2013.

[3] S. Li, A. Karatzoglou, and C. Gentile, "Collaborative filtering bandits," in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016, pp. 539–548.

[4] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult, "Toward open set recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 7, pp. 1757–1772, 2012.

[5] L. P. Jain, W. J. Scheirer, and T. E. Boult, "Multi-class open set recognition using probability of inclusion," in *European Conference on Computer Vision*. Springer, 2014, pp. 393–409.

[6] W. J. Scheirer, L. P. Jain, and T. E. Boult, "Probability models for open set recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 11, pp. 2317–2324, 2014.

[7] Q. Da, Y. Yu, and Z.-H. Zhou, "Learning with augmented class by exploiting unlabeled data," in *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 2014.

[8] H. Liu, Z. Cao, M. Long, J. Wang, and Q. Yang, "Separate to adapt: Open set domain adaptation via progressive separation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2927–2936.

[9] D. Pritsos and E. Stamatatos, "Open set evaluation of web genre identification," *Language Resources and Evaluation*, vol. 52, no. 4, pp. 949–968, 2018.

[10] H. Zhang and V. M. Patel, "Sparse representation-based open set recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 8, pp. 1690–1696, 2016.

[11] A. Bendale and T. Boult, "Towards open world recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1893–1902.

[12] R. De Rosa, T. Mensink, and B. Caputo, "Online open world recognition," *arXiv preprint arXiv:1604.02275*, 2016.

[13] P. R. M. Júnior, R. M. de Souza, R. d. O. Werneck, B. V. Stein, D. V. Pazinato, W. R. de Almeida, O. A. Penatti, R. d. S. Torres, and A. Rocha, "Nearest neighbors distance ratio open-set classifier," *Machine Learning*, vol. 106, no. 3, pp. 359–386, 2017.

[14] T. Doan and J. Kalita, "Overcoming the challenge for text classification in the open world," in *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 2017, pp. 1–7.

[15] E. M. Rudd, L. P. Jain, W. J. Scheirer, and T. E. Boult, "The extreme value machine," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 3, pp. 762–768, 2017.

[16] E. Vignotto and S. Engelke, "Extreme value theory for open set classification-gpd and gev classifiers," *arXiv preprint arXiv:1808.09902*, 2018.

[17] Z. Ge, S. Demyanov, Z. Chen, and R. Garnavi, "Generative openmax for multi-class open set classification," *arXiv preprint arXiv:1707.07418*, 2017.

[18] L. Neal, M. Olson, X. Fern, W.-K. Wong, and F. Li, "Open set learning with counterfactual images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 613–628.

[19] I. Jo, J. Kim, H. Kang, Y.-D. Kim, and S. Choi, "Open set recognition by regularising classifier with fake data generated by generative adversarial networks," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 2686–2690.

[20] A. R. Dhamija, M. Günther, and T. Boult, "Reducing network agnostophobia," in *Advances in Neural Information Processing Systems*, 2018, pp. 9157–9168.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in neural information processing systems*, 2015, pp. 91–99.

[22] S. Zhang, R. Benenson, M. Omran, J. Hosang, and B. Schiele, "Towards reaching human performance in pedestrian detection," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 973–986, 2017.

[23] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *European conference on computer vision*. Springer, 2016, pp. 21–37.

[24] A. Bendale and T. E. Boult, "Towards open set deep networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1563–1572.

[25] R. Yoshihashi, W. Shao, R. Kawakami, S. You, M. Iida, and T. Naemura, "Classification-reconstruction learning for open-set recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 4016–4025.

[26] D. O. Cardoso, F. França, and J. Gama, "A bounded neural network for open set recognition," in *2015 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2015, pp. 1–7.

[27] P. R. M. Júnior, T. E. Boult, J. Wainer, and A. Rocha, "Specialized support vector machines for open-set recognition," *arXiv preprint arXiv:1606.03802*, 2016.

[28] M. D. Scherreik and B. D. Rigling, "Open set recognition for automatic target classification with rejection," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 2, pp. 632–642, 2016.

[29] P. Panareda Busto and J. Gall, "Open set domain adaptation," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 754–763.

[30] Y. Yu, W.-Y. Qu, N. Li, and Z. Guo, "Open-category classification by adversarial sample generation," *arXiv preprint arXiv:1705.08722*, 2017.

[31] Y. Yang, C. Hou, Y. Lang, D. Guan, D. Huang, and J. Xu, "Open-set human activity recognition based on micro-doppler signatures," *Pattern Recognition*, vol. 85, pp. 60–69, 2019.

[32] K. Saito, S. Yamamoto, Y. Ushiku, and T. Harada, "Open set domain adaptation by backpropagation," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 153–168.

[33] D. O. Cardoso, J. Gama, and F. M. França, "Weightless neural networks for open set recognition," *Machine Learning*, vol. 106, no. 9-10, pp. 1547–1567, 2017.

[34] L. Shu, H. Xu, and B. Liu, "Doc: Deep open classification of text documents," *arXiv preprint arXiv:1709.08716*, 2017.

[35] P. Oza and V. M. Patel, "C2ae: Class conditioned auto-encoder for open-set recognition," *arXiv preprint arXiv:1904.01198*, 2019.

[36] ——, "Deep cnn-based multi-task learning for open-set recognition," *arXiv preprint arXiv:1903.03161*, 2019.

[37] Q. Lian, W. Li, L. Chen, and L. Duan, "Known-class aware self-ensemble for open set domain adaptation," *arXiv preprint arXiv:1905.01068*, 2019.