Open Set Recognition through Unsupervised and Class-Distance Learning

Andrew Draganov
adraganov@exptechinc.com
Expedition Technology, Inc.
Herndon, Virginia

Carter Brown
cbrown@exptechinc.com
Expedition Technology, Inc.
Herndon, Virginia

Enrico Mattei
emattei@exptechinc.com
Expedition Technology, Inc.
Herndon, Virginia

Cass Dalton
cdalton@exptechinc.com
Expedition Technology
Herndon, Virginia

Jaspreet Ranjit
jr4fs@virginia.edu
University of Virginia
Charlottesville, Virginia

ABSTRACT

We present a novel semi-supervised framework for training classifiers and simultaneously detecting out-of-distribution inputs. We do this by training on an closed classification dataset and an auxiliary simulated-open dataset, which consists of examples from outside the closed set. Through unsupervised learning and incorporating a class-distance value for each known class, we can identify out-of-distribution RF devices with state-of-the-art accuracy. We define metrics for quantifying robustness in terms of both classification and Open Set Recognition (OSR). Finally, we discuss uncertainty estimation and calibrate our open set predictions so that they represent confidence.

CCS CONCEPTS

• **Computing methodologies** → **Anomaly detection**.

KEYWORDS

Out of distribution detection, neural networks, open set recognition, RF fingerprinting

ACM Reference Format:
Andrew Draganov, Carter Brown, Enrico Mattei, Cass Dalton, and Jaspreet Ranjit. 2020. Open Set Recognition through Unsupervised and Class-Distance Learning. In *WiSec '20: 13th ACM Conference on Security and Privacy in Wireless and Mobile Networks, July 08–10, 2020, Linz, Austria.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/xxxxxxx.xxxxxxx

1 INTRODUCTION

Current classification models are usually trained to classify a set of known classes represented in the training set. However, in an open world setting, data also contains unknown examples from out-of-distribution classes. This presents a challenge to existing models since this "open set" of classes will be classified incorrectly as one of the known classes, often with a high degree of certainty. In some operational settings, it is critical for classification models to have a notion of the unknown unknowns in order to enable autonomous systems and incremental learning of new classes.

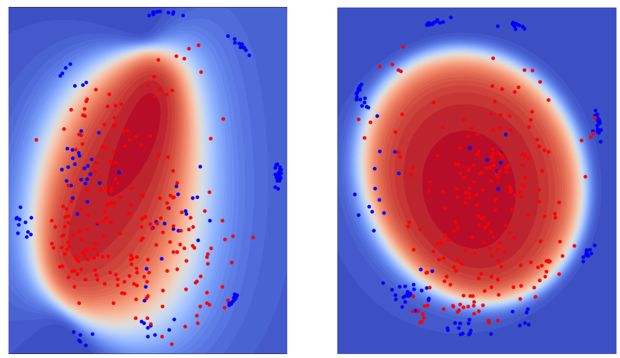


Figure 1: Effect of the unsupervised loss on the learned Open vs. Closed decision boundary in a latent space learned on a toy RF dataset. Blue points are closed, and red points are open. The decision boundary in the latent space is shaded from blue to red with respect to the openness score. On the left we show the result without the unsupervised loss, while the result using the unsupervised loss is shown on the right.

Numerous works have addressed the open set recognition problem using a diverse set of approaches such as the Compact Abating Probability (CAP) models [21]; exploiting the soft-max probabilities [8]; using auxiliary datasets for outlier exposure [6], [9]; generating confidence scores from a generative classifier [11]; class-conditioned auto-encoders [15]; and adversarial learning [4].

Contribution. In this work, we present a general approach for open-set recognition (OSR) which we call Unsupervised Class-Distance Learning (UCDL). We discuss new findings regarding OSR and uncertainty. We build on the idea introduced in [9] and use an auxiliary dataset, containing only open classes to promote learning a decision boundary between the closed and open sets. Specifically, we propose to jointly train two classifiers: one classifier is trained to solve the standard classification task (the known classes), while the second binary classifier is trained in a supervised fashion to classify closed vs open samples using both learned features and

Approved for Public Release, Distribution Unlimited

ACM ISBN 978-1-4503-XXXX-X/20/07…$15.00
https://doi.org/10.1145/xxxxxxx.xxxxxxx

their distances to the known classes. This enforces a partition of the feature space into closed and open regions. In addition, we propose to use a task-dependent unsupervised loss to encourage the model to learn general representations that can be leveraged for incremental learning of novel classes. We compare our model to traditional approaches, and report results on computer vision benchmarks as well as on a dataset composed of Radio Frequency (RF) emissions of Internet of Things (IoT) devices, made available by the DARPA Radio Frequency Machine Learning Systems (RFMLS) program.

2 RELATED WORK

There is a significant amount of existing literature regarding the open-set recognition problem. We highlight two major groups of methods and describe ours with respect to them.

Statistical methods fit the probability distribution of closed samples in a learned space to quantify the likelihood that a novel sample came from any of the known classes. These have found success through leveraging extreme value theory, which highlights the long-tailed nature of these closed-open distributions. The most notable of these approaches is OpenMax [2], which fits a Weibull distribution on the correctly classified softmax outputs. Similarly, Class Anchor Clustering [13] encourages clusters to form in the classifier outputs, which further separates the open samples from the closed ones. Another example is from [8], where they similarly look at the classification confidence distribution for closed vs. open samples.

Subspace methods learn the subspaces occupied by closed and open samples in a feature space. Network Agnostophobia [6] leverages the fact that open samples land near the origin in feature space and encourages this behavior on a sample dataset of open classes, showing that it transfers appropriately onto truly novel data. A disadvantage of this approach is that separability of open-set classes is not possible. In [19], a self-supervised domain adaptation method was used to promote clustering in feature space and to promote a better understanding of the closed vs. open subspace. The approach presented in [22] finds atypical samples across all closed classes and use those to represent the open set.

Quantifying uncertainty allows a model to predict a statistically-meaningful probability that a given input comes from an open set distribution or not. Platt Scaling [16] and its applications to other supervised models [14] including modern neural networks [7], provides a method to calibrate the outputs of a neural network's softmax to produce meaningful confidence scores. Furthermore, a model can learn to predict Aleatoric and Epistemic uncertainties [10] for a specific input and for the model itself. Finally, Smoothed Classifiers [5] and [20] provide methods to certify radii of robustness to Gaussian noising of the input-space for a given model.

Our model, UCDL, consolidates statistical and subspace approaches by providing both the feature vector as well as the distances to class centroids to an open set recognition module. This allows the network to learn the open/closed subspaces as well as the class-distance decision boundaries. Furthermore, by providing all of this in a single learned framework, we have the opportunity to quantify OSR uncertainty.

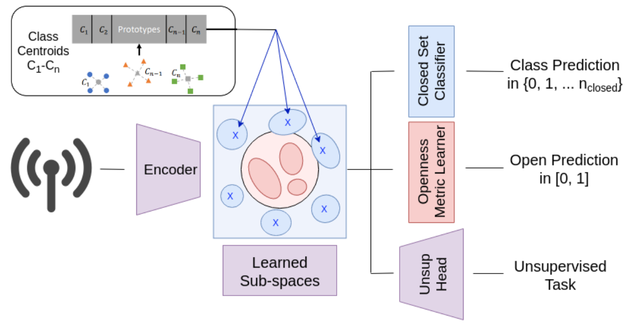


Figure 2: Diagram of our UCDL method with a feature extractor, learned latent space, and three task-specific network heads. Closed-set classification accomplished via a classification head that takes as input both the latent features from the network encoder and the features' distances to all the class centroids. The open set prediction branch takes as input both the latent features as well as the pairwise distance between each the data samples and all class prototypes. To encourage better feature extraction, a task-specific unsupervised task is incorporated as a third branch in the network.

3 APPROACH

3.1 RF Fingerprints

Here, we provide a summary of our approach on the RF datasets. The reader is referred to [12] for further details. Ideally, the signal transmitted by a wireless device can be expressed as

$$x(t) = Re\Bigg(A\Big(h_{re} * a_{re}(t)\cos(\omega(t)) + jh_{im} * a_{im}(t)\sin(\omega(t))\Big)e^{j2\pi f_c t}\Bigg), \tag{1}$$

where h_{re} and h_{im} are the impulse responses of the in-phase and quadrature reconstruction filters that make up the Digital-to-Analog Converter (DAC), $a(t)$ is the modulated digital signal amplitude, f_c is the carrier frequency, and A is the gain of the power amplifier. At the receiver, the signal can be modeled as $r(t) = x(t) * h_c + \eta(t)$, where h_c is the complex-valued impulse response of the propagation channel, $*$ denotes the convolution operator, and $\eta(t)$ is a Gaussian noise component $\sim \mathcal{CN}(0, \sigma_n^2)$.

It is well known that hardware imperfections cause the transmitted signal to deviate from its ideal representation, and this fact can be exploited for device identification, even when the device population contains many devices that are nominally identical. We exploit this physical difference between devices by recognizing that the DAC reconstruction filters are usually realized as relatively low-order analog filters. Therefore, we assume that the device-specific discriminative features occupy a low-dimensional subspace, according to the findings in [17].

3.2 Processing RF Data

Since digitized RF signals can be viewed as a time series, we employ sequence processing techniques to obtain a significant processing gain. We achieve this by exploiting the fact that a single device will transmit several bursts of data while attempting to gain access to a network. Here, a burst refers to the transmission of a data message. Let $X \in C^{M \times n}$ be the data matrix with rows representing M non-overlapping time windows of data from the same device, where each row $\boldsymbol{x}_i \in C^n$. From each $\boldsymbol{x}_i$ we obtain a corresponding latent representation $z_i \in C^d$, with $d < n$. The M latent representations should therefore be identical, and we denote this common latent vector as $\overline{z}$. We can then model $z_i | \boldsymbol{x}_i = \overline{z} + \boldsymbol{\eta}_i$, where $\boldsymbol{\eta}_i$ is a zero-mean Gaussian noise component. That is, the latent representation of each sequence element is a noisy observations of a sequence-level representation, $\overline{z}$, which we use for device classification.

We leverage this sequence-based processing scheme for unsupervised learning by constraining each sequence to contain only data from the same burst, so we process data one burst at a time. With this construction, we assume that any two signal bursts drawn from the unlabeled dataset belong to different devices with high likelihood, and we compute a contrastive loss between pairs of bursts as follows,

$$\mathcal{L}_{ij} = \|f(\boldsymbol{X}_i) - f(\boldsymbol{P}\boldsymbol{X}_i)\|_F^2 - \|f(\boldsymbol{X}_i) - f(\boldsymbol{X}_j)\|_F^2, \qquad (2)$$

where $f(\cdot)$ is the neural network encoder, $\boldsymbol{P}$ is a row permutation matrix, and $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$ are the data matrices for two different signal bursts.

3.3 Open Set Recognition Datasets and Metrics

In order to train the joint tasks of traditional classification and open set recognition, we split the data into three disjoint groups of classes - closed classes, sim-open classes, and true-open classes. The closed datasets have a class label c and openness label 0, while the open datasets have no class label and openness label 1. This data is easy to obtain, as many tasks have small labeled datasets and large unlabeled datasets, which we simply relabel as closed and open. We use only the closed classes to learn the traditional classification task but use both the closed and simulated-open classes to learn the OSR task. We then analyze generalization to novel devices on the true-open classes. We note that using the unlabeled dataset as sim-open may result in noisy labels, as some unlabeled examples may actually come from the closed set. This is can be overcome however, by enforcing that the openness predictions represent open/closed confidences, as is shown in the results section. We can also apply noisy learning techniques that exploit quantifiable feature consistency such as in [18].

Furthermore, we define metrics that fully describe the efficacy of an open set recognition model. With the introduction of out-of-distribution samples, it is no longer sufficient to simply return the accuracy on the closed set, as one could classify a sample correctly but mis-identify it as open. Therefore, we use the following metrics: the *traditional class accuracy* ($\text{ACC}^{\text{traditional}}$) represents the percent of closed samples that are correctly classified. The *amended class accuracy* ($\text{ACC}^{\text{amended}}$), however, indicates the percent of closed samples that are *both* correctly classified *and* identified as closed. The *false positive rate* (FPR) the percent of closed samples that are falsely identified as open. The *true positive rate* (TPR) the percent of true open samples that are correctly identified as open. Note that each metric is constrained to just one of the dataset splits. The $\text{ACC}^{\text{amended}}$, FPR, and TPR are calculated with respect to an openness threshold, which is a binary decision boundary on the network's openness predictions. We determine the optimal threshold using the train set, and then apply it onto the evaluation set. We posit that an effective OSR method must obtain both a high amended class accuracy, a high TPR, and a low FPR.

3.4 Network Architecture

Our network architecture is inspired by the realization that the open set recognition problem can be easily tied to the new developments in semi-supervised learning. Namely, we have a labeled closed dataset and plentiful unlabeled data that we may define as open. By assigning openness values onto the unlabeled data, we now have a way to simultaneously supervise open set recognition and semi-supervise closed set classification.

With this in mind, we construct our network to perform both traditional classification as well as open set recognition in one forward pass, training both tasks in a single learned space. We train on input batches that are half-closed and half-simulated-open. These pass through a convolutional feature extractor that extracts relevant latent information. To incorporate distance-based methods, we obtain centroids for each class in this feature space and ensure that our feature vector corresponds to one of these centroids.

We provide these latent representations to a classifier that produces a softmax over the closed classes, an openness predictor that produces a scalar in the range of [0, 1], and an unsupervised head that encourages better feature extraction. In the case of the openness predictor, we feed it *both* the feature embedding as well as the distance of these features to every class centroid, so that it may perform both subspace and distance-based OSR.

During training, we do not let the feature extractor receive gradients from the open set recognition head. We found that supervising the feature extraction on the closed and simulated open samples results in heavy overfitting towards *only* the sim-open classes being recognized as open. For this reason, we use an unsupervised loss to encourage learning on the simulated open data without overfitting to the fact that it is open.

Lastly, we employ Platt scaling [16], to recalibrate the openness predictions as described in 4.2 to ensure that the openness scalars represent the network's open-set confidence.

4 RESULTS

4.1 RF Fingerprinting Datasets

We report results on RF fingerprinting for a baseline experiment, as well as for several experiment sweeps. The architecture here is a complex-valued convolutional network as proposed in [12]. We use a contrastive loss for the unsupervised task, where samples of the same class are encouraged to attract and samples from different classes are encouraged to repel. Data augmentation was applied during training as discussed in Section 4.3. The baseline for comparison contains 100 closed wifi classes with 200 bursts per class, 100 sim-open classes with 200 bursts per class, and 400 true-open classes with 30 examples per class. All RF experiments have these default dataset sizes unless a modification is otherwise specified.

In Table 1, we highlight our RF fingerprinting OSR results. The principal findings are that OSR performance improves with higher sim-open class sampling, but that this improvement may make

Experiment Group	Experiment Parameter	ACC$^{\text{traditional}}$	ACC$^{\text{amended}}$ at threshold 0.5	TPR	FPR
Baseline	N/A	**89.7%**	75.7%	81.5%	15.6%
With and without class distances and unsup. loss	Only class-distances	88.9%	74.5%	82.8%	17.2%
	Only unsup. loss	89.5%	76.8%	68.5%	15.0%
	Neither	88.3%	77.8%	66.4%	13.2%
Sweep over number of sim-open classes	20	86.4%	**81.7%**	63.5%	**7.9%**
	50	87.5%	80.4%	75.9%	10.8%
	200	85.6%	76.8%	**83.4%**	13.8%

Table 1: Multiple ablation studies against the baseline described in 4.1. These show that an improved separation for TPR – FPR may lead to a slightly lower ACC$^{\text{amended}}$, as the network begins to favor labeling samples as open. We also see the effect that the unsupervised and class-distance vectors have on both the open and closed set performance. Obtaining a balance between the TPR and FPR requires a roughly even number of classes in the closed and sim-open datasets.

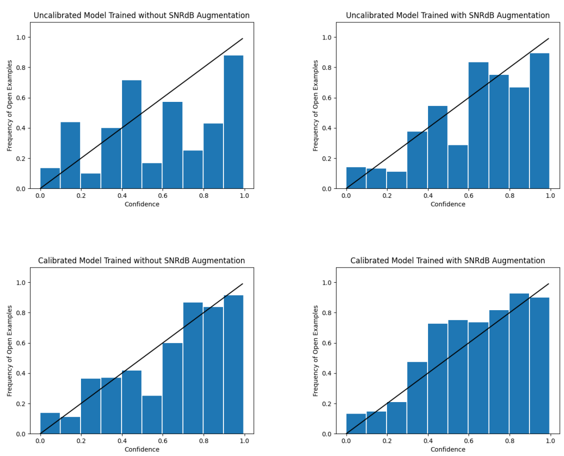


Figure 3: Top row: uncalibrated reliability diagrams of openness predictions for UCDL models without and with simulated SNR augmentation as described in 4.3. Bottom row: reliability diagrams for openness predictions after Platt scaling calibration as described in 4.2. Left column: UCDL model trained without SNR augmentation. Right column: UCDL model trained with SNR augmentation. The x-axes correspond to output confidence and the y-axis corresponds to what proportion of the samples were open.

the network more trigger-happy to call samples open. We also see that both the unsupervised and the class-distance are helpful for improving open set recognition performance.

Finally, we calibrate our openness scores with Platt Scaling on a model trained on 30 examples per each of the 100 closed classes and 200 simulated open classes with 400 examples each. We find that Platt Scaling qualitatively makes the openness reliability diagrams more diagonal and therefore more interpretable as confidences. UCDL trained with simulated SNR augmentation and Platt calibration is less overconfident with its openness scores. SNR augmentation and Platt calibration also each reduce the openness binary cross-entropy as shown in Table 2.

Calibration	No SNR Augmentation	SNR Augmentation
Before Platt Scaling	0.683	0.571
After Platt Scaling	0.441	**0.437**

Table 2: Binary cross-entropy loss of predicted openness scores before and after Platt Scaling, with and without SNR augmentation during model training.

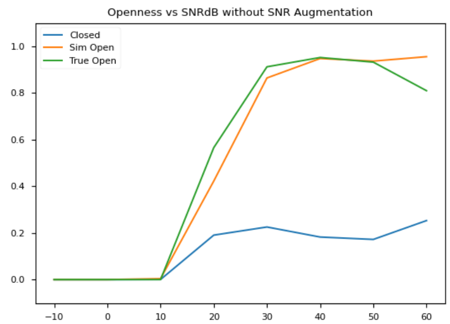
4.2 Calibrating Openness

As shown in [7], modern neural networks are not well-calibrated, i.e. the outputs of their final classification layers tend to be severely overconfident and need to be calibrated. Hence, we verify similar results for our OSR RF models and calibrate the openness scores using Platt Scaling [16]. We find that while the reliability diagrams of the openness scores from the OSR model aren't perfectly diagonal, they are not as severely overconfident as found in [7].

Figure 3 shows the uncalibrated reliability diagrams from a sample of the eval set such that the number of open samples is the same as the number of closed samples, and the number of open samples is half simulated open and half true open. We find that the model trained with SNR data augmentation as described in Section 4.3 has a qualitatively better reliability diagram that is more diagonal and has more monotonicity. We next apply Platt Scaling [16] to calibrate the openness scores of both models. Qualitatively, we verify that the Platt Scaling improves the diagonalization of the reliability diagram. The model trained with SNR augmentation obtains the least overconfidence after calibration. Quantitatively, Platt Scaling reduces the binary cross-entropy of openness predictions, and the model trained with SNR augmentation also has a lower binary cross-entropy in Table 2.

4.3 Robustness of RF Openness

Data augmentation is frequently used during training to improve inference accuracy and robustness to perturbations of the input space. In the RF domain, works including [1], [12], [3] have demonstrated the importance of adding data augmentation for inference-time accuracy. We show these augmentations also help with OSR accuracy. Specifically, we simulate noise in the data, as described in Section 3.1, by adding white Gaussian noise (AWGN). We estimate the power of the signal as the total power of the non-augmented signal to estimate a signal-to-noise (SNR) ratio of the output signal. This SNR representation allows better interpretation of results than



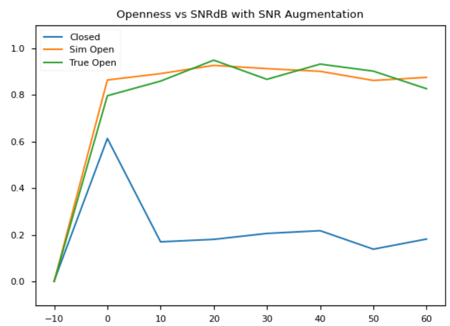


Figure 4: Uncalibrated smoothed openness scores for examples from closed, simulated open, and true open distributions. Models trained with simulated SNR augmentation (top plot) have robust openness scores over a wider range of SNRs than their non-augmented counterparts (bottom plot). For each value of simulated SNR as described in 4.3, samples from the three distributions are applied with multiple instantiations of artificial SNR, and each sample's openness scores is averaged over the set of applied simulated SNRs. On the top are the results of the OSR model with no SNR augmentation during training. Results for the OSR model with SNR augmentation during training are on the bottom.

a noise variance representation for extending robustness results a la [5] and [20].

We perform an A/B test on the experiment of 30 closed examples per class and 200 sim-open classes from 1. To one model we apply only center frequency offset and channel augmentations, as described in [12], [3], to our OSR RF model with unsupervised loss, while to the other model we also augment with AWGN to produce an artificial estimated SNR within the range of 5dB to 50dB with a uniform distribution in dB-space.

Figure 4 shows that AWGN augmentation leads to openness scores that are robust over a wider range of SNR values. Once the added noise pushes SNR below a level that can be handled by the OSR model, 30dB in the non-augmented case and 0dB in the augmented case, the openness scores have a strong bias towards 0. This is counterintuitive, for very noisy data would be expected either to tend towards 0.5 or towards 1 from an entropic standpoint. However, since we don't have the model learn to connect entropy of the input and output spaces, there's no reason we would expect those entropies to be correlated, and the model is likely correlating with an imbalance in the training sets.

5 CONCLUSION

In this work we discussed existing methods for open set recognition and proposed a cohesive approach to consolidate them in a single learned subspace. By re-framing unlabeled data as open, we supervise an open set recognition network while also encouraging robust feature extraction with class centroids and an unsupervised loss. Furthermore, by learning the open subspace and its distance to every known class, we can qualitatively and quantitatively analyze open set uncertainty.

6 ACKNOWLEDGMENTS

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA). The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

REFERENCES

[1] Ioannis Agadakos, Nikolaos Agadakos, Jason Polakis, and Mohamed R. Amer. 2019. Deep Complex Networks for Protocol-Agnostic Radio Frequency Device Fingerprinting in the Wild. *arXiv e-prints*, Article arXiv:1909.08703 (Sept. 2019), arXiv:1909.08703 pages. arXiv:cs.CR/1909.08703

[2] Abhijit Bendale and Terrance E. Boult. 2016. Towards Open Set Deep Networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 1563–1572.

[3] Metehan Cekic, Soorya Gopalakrishnan, and Upamanyu Madhow. 2020. Robust Wireless Fingerprinting: Generalizing Across Space and Time. *ArXiv* abs/2002.10791 (2020).

[4] Jiefeng Chen, Yixuan Li, Xi Wu, Yingyu Liang, and Somesh Jha. 2020. Robust Out-of-distribution Detection in Neural Networks. *arXiv preprint arXiv:2003.09711* (2020).

[5] Jeremy Cohen, Elan Rosenfeld, and Zico Kolter. 2019. Certified Adversarial Robustness via Randomized Smoothing. In *Proceedings of the 36th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Kamalika Chaudhuri and Ruslan Salakhutdinov (Eds.), Vol. 97. PMLR, Long Beach, California, USA, 1310–1320. http://proceedings.mlr.press/v97/cohen19c.html

[6] Akshay Raj Dhamija, Manuel Günther, and Terrance Boult. 2018. Reducing Network Agnostophobia. In *NIPS 2018: The 32nd Annual Conference on Neural Information Processing Systems*. 9157–9168.

[7] Chuan Guo, Geoff Pleiss, Yu Sun, and Killian Q. Weinberger. 2017. On Calibration of Modern Neural Networks. In *Proceedings of the 34th International Conference on Machine Learning*. ICML, 1321–1330.

[8] Dan Hendrycks and Kevin Gimpel. 2017. A Baseline for Detecting Misclassified and Out-of-Distribution Examples in Neural Networks. In *ICLR 2017 : International Conference on Learning Representations 2017*.

[9] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. 2019. Deep Anomaly Detection with Outlier Exposure. In *ICLR 2019 : 7th International Conference on Learning Representations*.

[10] Alex Kendall and Yarin Gal. 2017. What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision?. In *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, 4-9 December 2017, Long Beach, CA, USA*, Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (Eds.). 5574–5584.

[11] Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. 2018. A Simple Unified Framework for Detecting Out-of-Distribution Samples and Adversarial Attacks. In *NIPS 2018: The 32nd Annual Conference on Neural Information Processing Systems*. 7167–7177.

[12] Enrico Mattei, Cass Dalton, Andrew Draganov, Brent Marin, Michael Tinston, Greg Harrison, Bob Smarrelli, and Marc Harlacher. 2019. Feature Learning for Enhanced Security in the Internet of Things. In *2019 IEEE Global Conference on Signal and Information Processing*.

[13] Dimity Miller, Niko Sünderhauf, Michael Milford, and Feras Dayoub. 2020. Class Anchor Clustering: a Distance-based Loss for Training Open Set Classifiers. arXiv:cs.CV/2004.02434

[14] Alexandru Niculescu-Mizil and Rich Caruana. 2005. Predicting Good Probabilities with Supervised Learning. In *Proceedings of the 22nd International Conference on Machine Learning*. ICML, 625–632.

[15] Poojan Oza and Vishal M. Patel. 2019. C2AE: Class Conditioned Auto-Encoder for Open-Set Recognition. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2307–2316.

[16] John C. Platt. 1999. Probabilistic Outputs for Support Vector Machines and Comparisons to Regularized Likelihood Methods. In *Advances in Large Margin Classifiers*. MIT Press, 61–74.

[17] Adam C Polak, Sepideh Dolatshahi, and Dennis L Goeckel. 2011. Identifying wireless users via transmitter imperfections. *IEEE Journal on selected areas in communications* 29, 7 (2011), 1469–1479.

[18] Scott Reed, Honglak Lee, Dragomir Anguelov, Christian Szegedy, Dumitru Erhan, and Andrew Rabinovich. 2015. Training Deep Neural Networks on Noisy Labels with Bootstrapping. In *ICLR 2015*.

[19] Kuniaki Saito, Donghyun Kim, Stan Sclaroff, and Kate Saenko. 2020. Universal Domain Adaptation through Self Supervision. arXiv:cs.CV/2002.07953

[20] Hadi Salman, Greg Yang, Jerry Li, Pengchuan Zhang, Huan Zhang, Ilya Razenshteyn, and Sébastien Bubeck. 2019. Provably Robust Deep Learning via Adversarially Trained Smoothed Classifiers. In *33rd Conference on Neural Information Processing Systems (NeurIPS 2019)*.

[21] W. J. Scheirer, A. de Rezende Rocha, A. Sapkota, and T. E. Boult. 2013. Toward Open Set Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 7 (2013), 1757–1772.

[22] Patrick Schlachter, Yiwen Liao, and Bin Yang. 2019. Open-Set Recognition Using Intra-Class Splitting. *2019 27th European Signal Processing Conference (EUSIPCO)* (Sep 2019). https://doi.org/10.23919/eusipco.2019.8902738