Anomaly Detection for OOD and Novel Category Detection

Thomas G. Dietterich, Oregon State University

tgd@cs.orst.edu

@tdietterich

Grad students: Andrew Emmott, Si Liu, Risheek Garrepalli, Alex Guyer, Dan Hendrycks
Faculty members: Alan Fern, Debashis Mondal

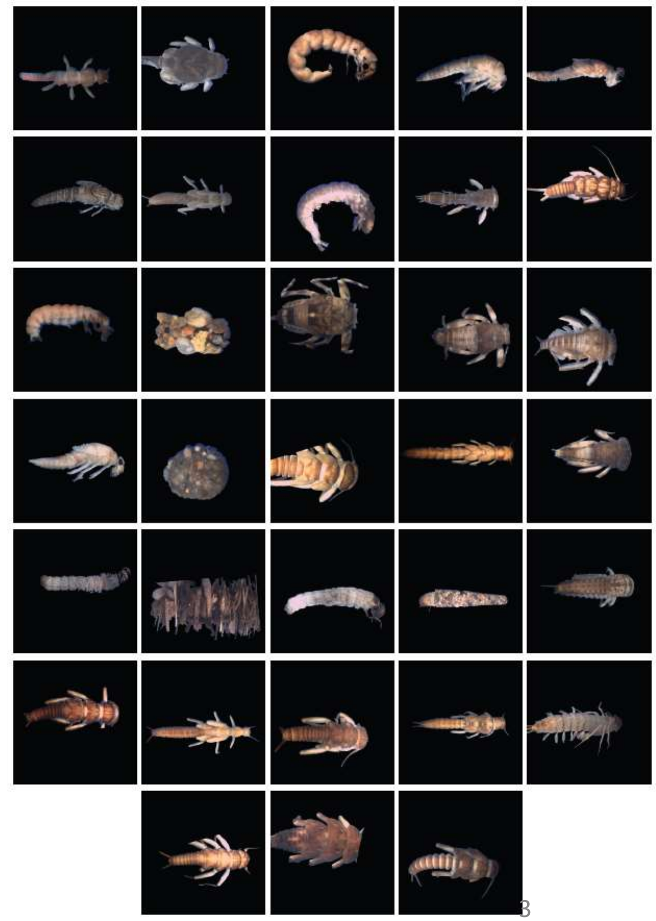
Thank you Dan Hendrycks and Balaji Lakshminarayanan for advice and suggestions

Motivating Example: Automated Counting of Freshwater Macroinvertebrates

- Goal: Assess the health of freshwater streams
- Method:
  - Collect specimens via kicknet
  - Photograph in the lab
  - Classify to genus and species
- BugID Project
  - 54 classes of interest to the EPA
  - accuracy $\approx 90\%$
  - Larios, N., Soran, B., Shapiro, L., Martínez-Muños, G., Lin, J., Dietterich, T. G. (2010). **Haar Random Forest Features and SVM Spatial Matching Kernel for Stonefly Species Identification.** *IEEE International Conference on Pattern Recognition (ICPR-2010)*.
  - Lin, J., Larios, N., Lytle, D., Moldenke, A., Paasch, R., Shapiro, L., Todorovic, S., Dietterich, T. (2011). **Fine-Grained Recognition for Arthropod Field Surveys: Three Image Collections**. *First Workshop on Fine-Grained Visual Categorization (CVPR-2011)*
  - Lytle, D. A., Martínez-Muñoz, G., Zhang, W., Larios, N., Shapiro, L., Paasch, R., Moldenke, A., Mortensen, E. A., Todorovic, S., Dietterich, T. G. (2010). **Automated processing and identification of benthic invertebrate samples.** *Journal of the North American Benthological Society*, 29(3), 867-874.

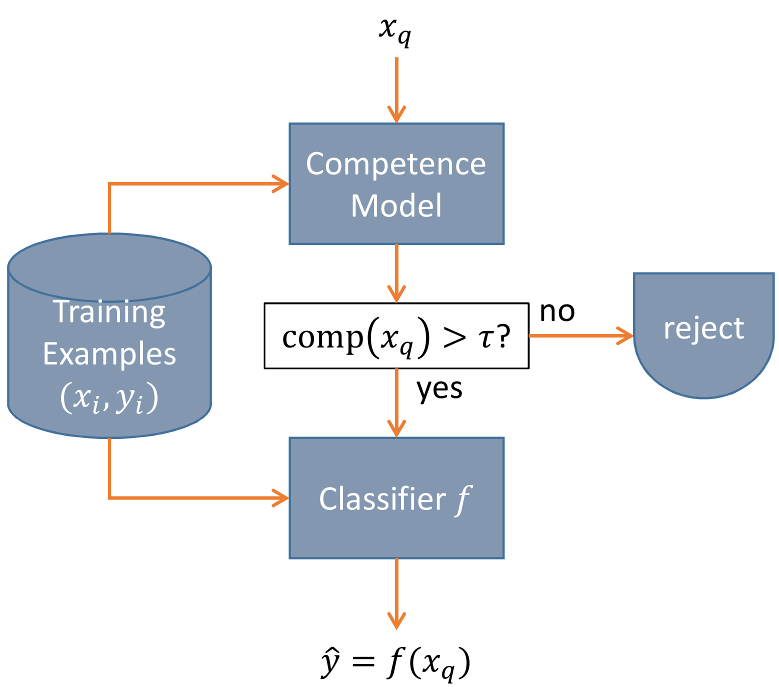
Problem: There are $\approx 76{,}000$ species of freshwater insects worldwide

- 1,200 species in US
- Field samples may contain other things
  - leaves
  - trash
- Simple estimate of equal error rate for novel classes vs. the 54 classes was 20% (in 2011)
  - classifier is not usable without addressing the novel class problem
- Open Category Problem



Open Set/OOD Classification Problem

- Arises in any application of classification in an open world
  - Novel obstacles in self-driving cars
  - Novel diseases in medical imaging
  - Novel products in online marketplaces
  - Novel cyber attacks
- Claim: Every deployed ML classifier should include a competence model that can detect when new queries are far from the training data



Two Problem Formulations: OOD and Open Category

Out-of-Distribution Problem

- Training:
  - Data: $(x_1, y_1), \ldots, (x_N, y_N)$ drawn from D_0
  - $y_i \in \{1, \ldots, K\}$
- Testing:
  - Data: Mixture D_m of data from D_0 and D_a
  - $(x, y) \sim D_a$ belong to a **different data set**
- Goal:
  - Given a query x_q, does it belong to D_a or D_0?
  - If from D_a, REJECT as alien
  - Else classify using a classifier trained on D_0 data

Novel Category / Open Set Problem

- Training:
  - Data: $(x_1, y_1), \ldots, (x_N, y_N)$ drawn from D_0
  - $y_i \in \{1, \ldots, K\}$
- Testing:
  - Data: Mixture D_m of data from D_0 and D_a
  - $(x, y) \sim D_a$ belong to **new classes not seen during training ("alien categories")**
- Goal:
  - Given a query x_q, does it belong to D_a or D_0?
  - If from D_a, REJECT as alien
  - Else classify using a classifier trained on D_0 data

OOD and Novel Category Metrics

- AUROC: Area under the ROC curve for the binary decision
  - OOD: Domain A vs Domain B
  - Novel Category: Known vs Unknown
  - We treat the anomalies as the positive class
- Detection rate at fixed false alarm rate. Maximize TPR@10%FPR
  - Maximize correct OOD/Novel Category detections subject to a constraint that the false positive rate is ≤ 0.10.
- False alarm rate at fixed missed alarm rate: Minimize FPR@95%TPR
  - Detect 95% of OOD/Novel Category examples while minimizing false positives
  - Most relevant to AI Safety and Trustworthy Systems

Outline

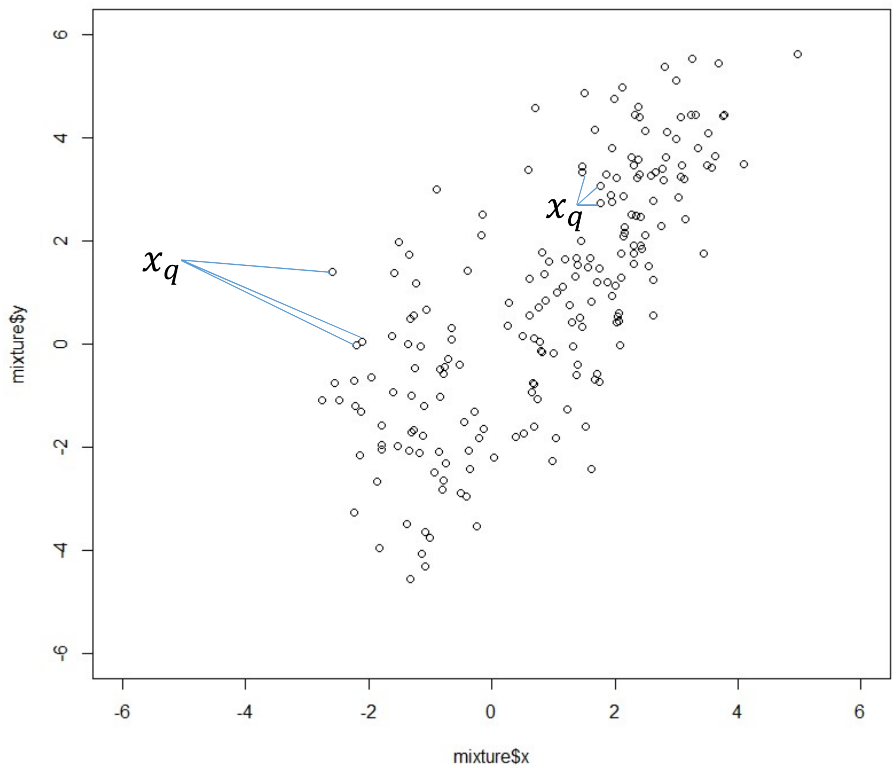
- Theoretical Approaches to Anomaly Detection
- Deep Anomaly Detection in Computer Vision
- Anomaly Detection based on Supervised Classifier Logit Scores

Theoretical Approaches to Anomaly Detection

- **Distance-Based Methods**
  - Anomaly score
    $$A(x_q) = \min_{x \in D} \|x_q - x\|$$
- **Density Estimation Methods**
  - Surprise: $A(x_q) = -\log P_D(x_q)$
  - Model the joint distribution $P_D(x)$ of the input data points $x_1, \ldots \in D$
- **Quantile Methods**
  - Find a smooth function f such that $\{x : f(x) \geq 0\}$ contains $1 - \alpha$ of the training data
  - Anomaly score $A(x) = -f(x)$
- **Reconstruction Methods**
  - Train an auto-encoder: $x \approx D(E(x))$, where E is the encoder and D is the decoder
  - Anomaly score
    $$A(x_q) = \left\| x_q - D\left(E(x_q)\right) \right\|$$

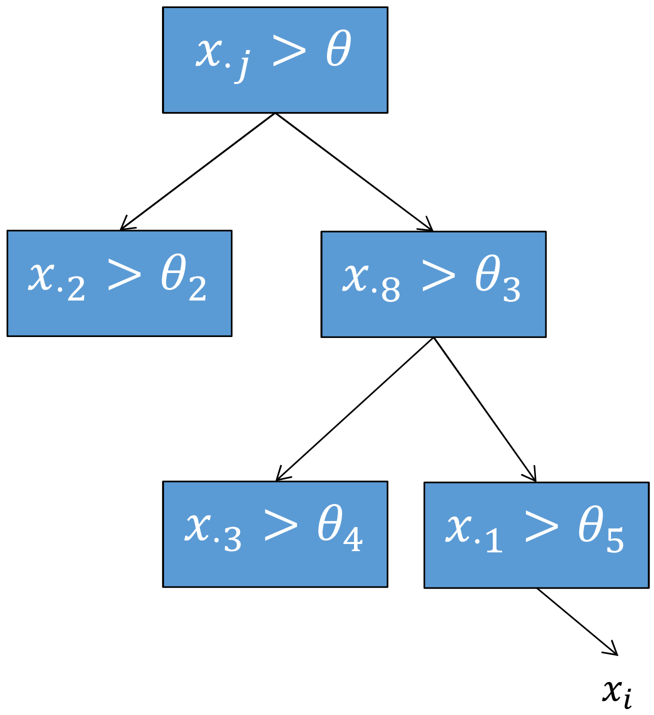
Approach 1: Distance-Based Methods

- Define a distance $d(x_i, x_j)$
- $A(x_q) = \min_{x \in D} d(x_q, x)$
- Requires a good distance metric



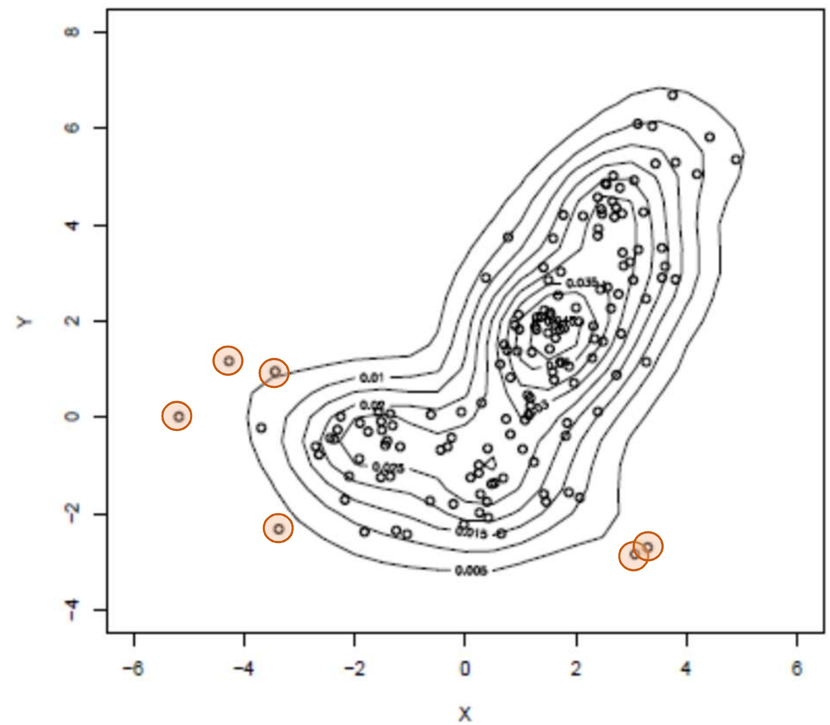
Isolation Forest [Liu, Ting, Zhou, 2011]

- Approximates L1 Distance
  - (Guha, et al., ICML 2016)
- Construct a fully random binary tree
  - choose attribute j at random
  - choose splitting threshold θ uniformly from $\left[\min\left(x_{\cdot j}\right), \max\left(x_{\cdot j}\right)\right]$
  - until every data point is in its own leaf
  - let $d(x_i)$ be the depth of point x_i
- repeat L times
  - let $\bar{d}(x_i)$ be the average depth of x_i
  - $A(x_i) = 2^{-\left(\frac{\bar{d}(x_i)}{r(x_i)}\right)}$
    - $r(x_i)$ is the expected depth



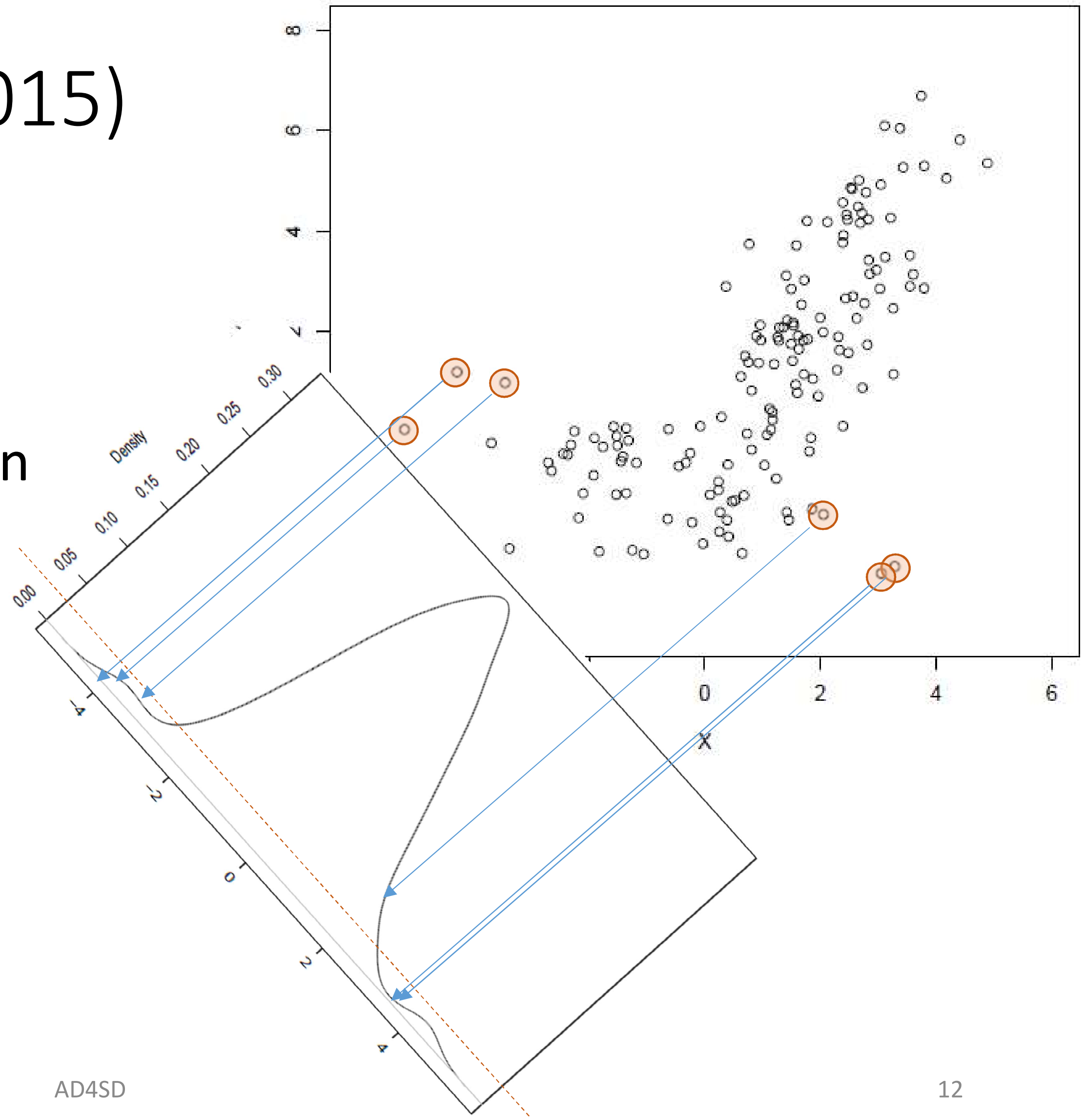
Approach 2: Density Estimation

- Given a data set $\{x_1, \dots, x_N\}$ where $x_i \in \mathbb{R}^d$
- We assume the data have been drawn iid from an unknown probability density: $x_i \sim P(x_i)$
- Goal: Estimate P
- Anomaly Score: $A(x_q) = -\log P(x_q)$
  - "surprisal" from information theory
- Why density estimation?
  - Gives a more global view by combining distances to all data points



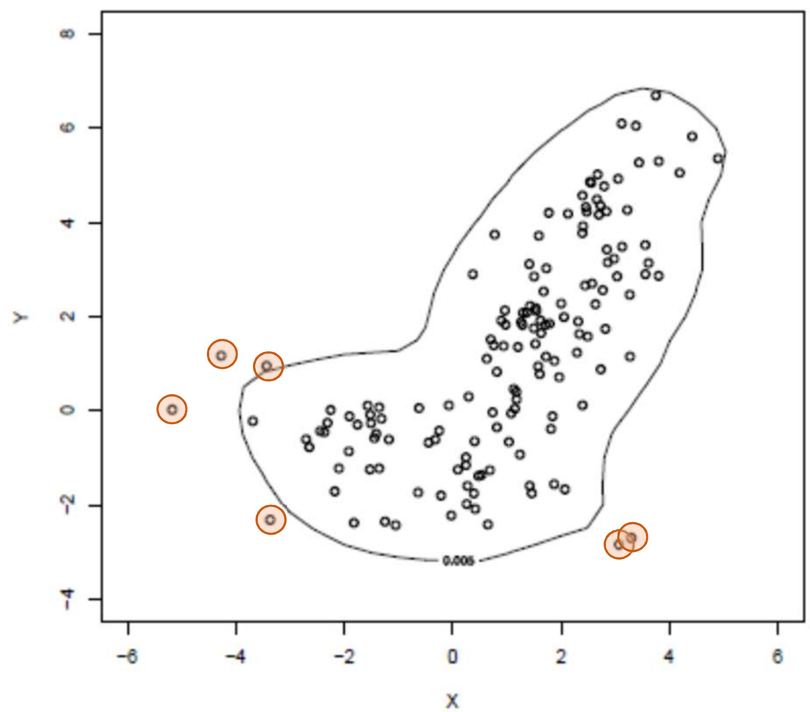
Example: LODA (Pevny, 2015)

- Introduce random projections Π_l into 1-dimensional space
- Fit a density estimator $P_l\left(\Pi_l(x)\right)$ in each 1-d space
- $A(x) = \frac{1}{L}\sum_{l=1}^{L} -\log P_l\left(\Pi_l\left(x_q\right)\right)$



Approach 3: Quantile Methods

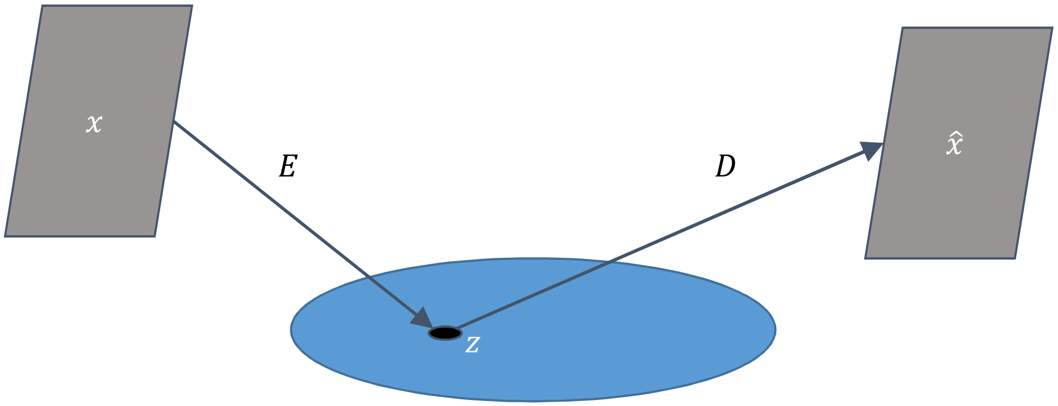
- Vapnik's principle: We only need to estimate the "decision boundary" between nominal and anomalous
- Surround the data by a function f that captures $1 - \epsilon$ of the training data
  - One-Class Support Vector Machine (OCSVM)
    - f is a hyperplane in "kernel space"
  - Support Vector Data Description (SVDD)
    - f is a sphere is "kernel space"
- Issue
  - Need to choose ϵ at learning time rather than run time



Approach 4: Reconstruction Methods

Autoencoders

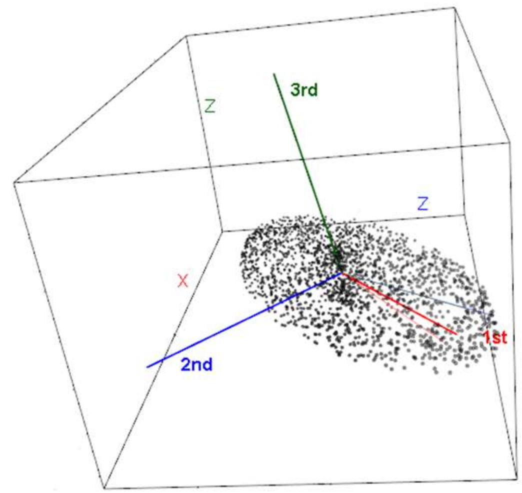
- Encoder: $z = E(x)$
- Decoder: $\hat{x} = D(z)$



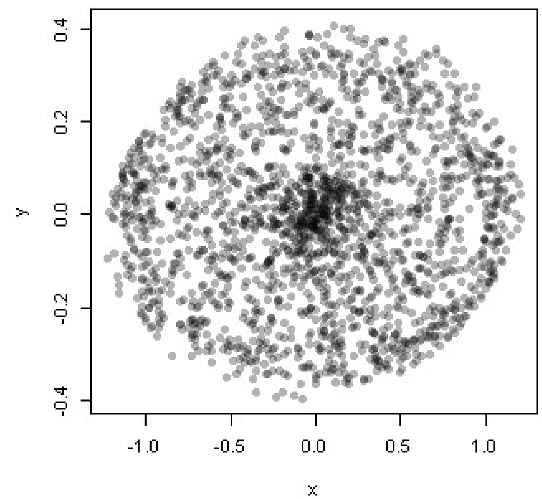
Linear Autoencoder == Principal Component Analysis

- PCA:
  - Let the input dimension be d
  - Choose a latent dimension ℓ
  - Find the $d \times \ell$ matrix W that minimizes the squared reconstruction error
  - $\min_W \sum_i \|x_i - WW^\top x_i\|^2$
    - This can be done using the Singular Value Decomposition
    - It can also be viewed as fitting a multi-variate Gaussian to the data and then keeping only the ℓ dimensions of highest variance

PCA applied to an ellipsoidically shaped point cloud



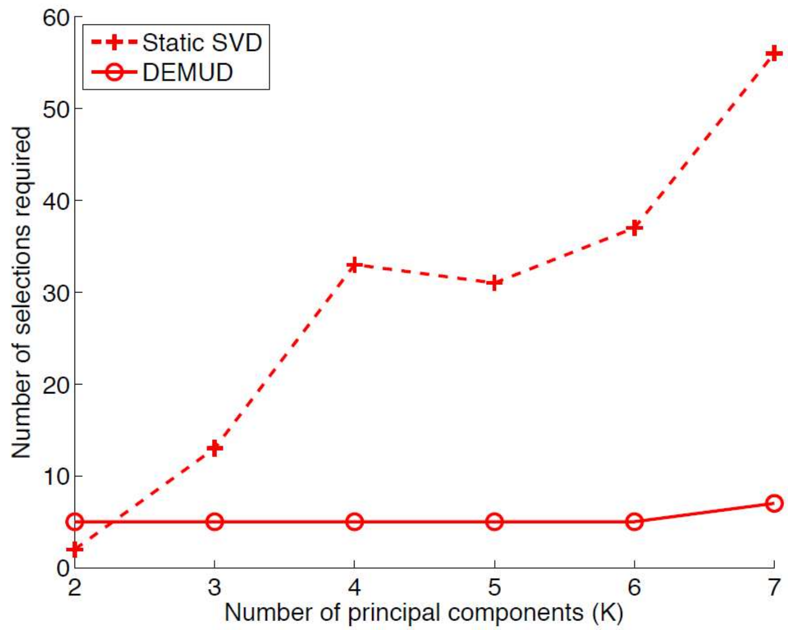
Projection against 1st and 2nd eigen vector



https://www.joyofdata.de/blog/illustration-of-principal-component-analysis-pca/

Application: Finding Unusual Chemical Spectra

- NASA Mars Science Laboratory ChemCam instrument
  - Collects 6144 spectral bands on rock samples from 7m distance using laser stimulation
  - Goal: active learning to find interesting spectra
  - DEMUD
    - Incremental PCA applied to samples one at a time
    - Fit only to the samples labeled as "uninteresting" by the user
    - Show the user the most un-uninteresting sample (sample with highest PCA reconstruction error)
  - Rapidly discovers interesting samples
  - Wagstaff, et al. (2013)



(a) Effort required to discover magnesite.

Benchmarking Study

[Andrew Emmott, 2015, 2020]

- Distance-Based Methods
  - k-NN: Mean distance to k-nearest neighbors
  - LOF: Local Outlier Factor (Breunig, et al., 2000)
  - ABOD: kNN Angle-Based Outlier Detector (Kriegel, et al., 2008)
  - IFOR: Isolation Forest (Liu, et al., 2008)
- Density-Based Approaches
  - RKDE: Robust Kernel Density Estimation (Kim & Scott, 2008)
  - EGMM: Ensemble Gaussian Mixture Model (our group)
  - LODA: Lightweight Online Detector of Anomalies (Pevny, 2016)
- Quantile-Based Methods
  - OCSVM: One-class SVM (Schoelkopf, et al., 1999)
  - SVDD: Support Vector Data Description (Tax & Duin, 2004)

Benchmarking Methodology

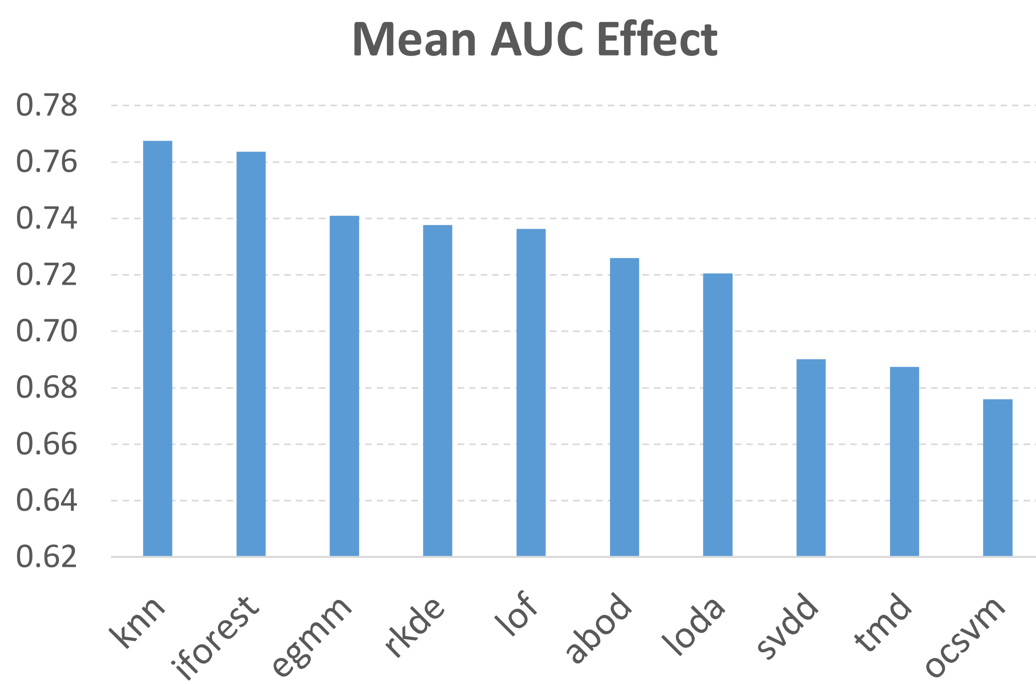
- Select 19 data sets from UC Irvine repository
- Choose one or more classes to be “anomalies”; the rest are “nominals”
- Manipulate
  - Relative frequency
  - Point difficulty
  - Irrelevant features
  - Clusteredness
- 20 replicates of each configuration
- Result: 11,888 Non-trivial Benchmark Datasets

Analysis of Variance

- Linear ANOVA
  - $\log \frac{AUC}{1 - AU} \sim rf + pd + cl + ir + pset + algo$
    - rf: relative frequency
    - pd: point difficulty
    - cl: normalized clusteredness
    - ir: irrelevant features
    - mset: "Parent" set
    - algo: anomaly detection algorithm
- Assess the *algo* effect while controlling for all other factors
- AUC: area under the ROC curve for the nominal vs. anomaly binary decision

Benchmarking Study Results

- 19 UCI Datasets
- 9 Leading "feature-based" algorithms
- 11,888 non-trivial benchmark datasets
- Mean AUC effect for "nominal" vs. "anomaly" decisions
    - Controlling for
        - Parent data set
        - Difficulty of individual queries
        - Fraction of anomalies
        - Irrelevant features
        - Clusteredness of anomalies
- Baseline method: Distance to nominal mean ("tmd")
- Best methods: K-nearest neighbors and Isolation Forest
- Worst methods: Kernel-based OCSVM and SVDD



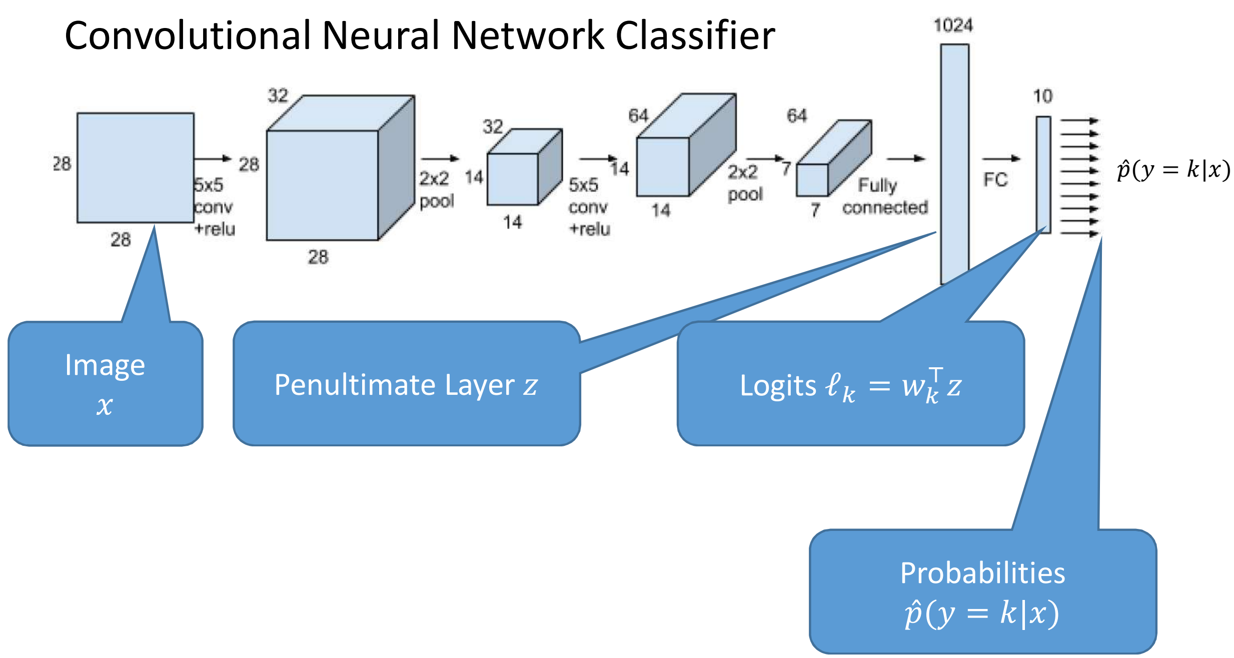
Outline

- Theoretical Approaches to Anomaly Detection
- Deep Anomaly Detection in Computer Vision
- Anomaly Detection based on Supervised Classifier Logit Scores

Deep Anomaly Detection in Image Classification

- Input image x
- Network backbone, also called the "encoder": $z = E(x)$
- Latent representation z
- Logits $\ell_k = w_k^\top z$
- Predicted probabilities

$$\hat{p}(y = k|x) = \frac{\exp \ell_k(z)}{\sum_{k'} \exp \ell_{k'}(z)}$$



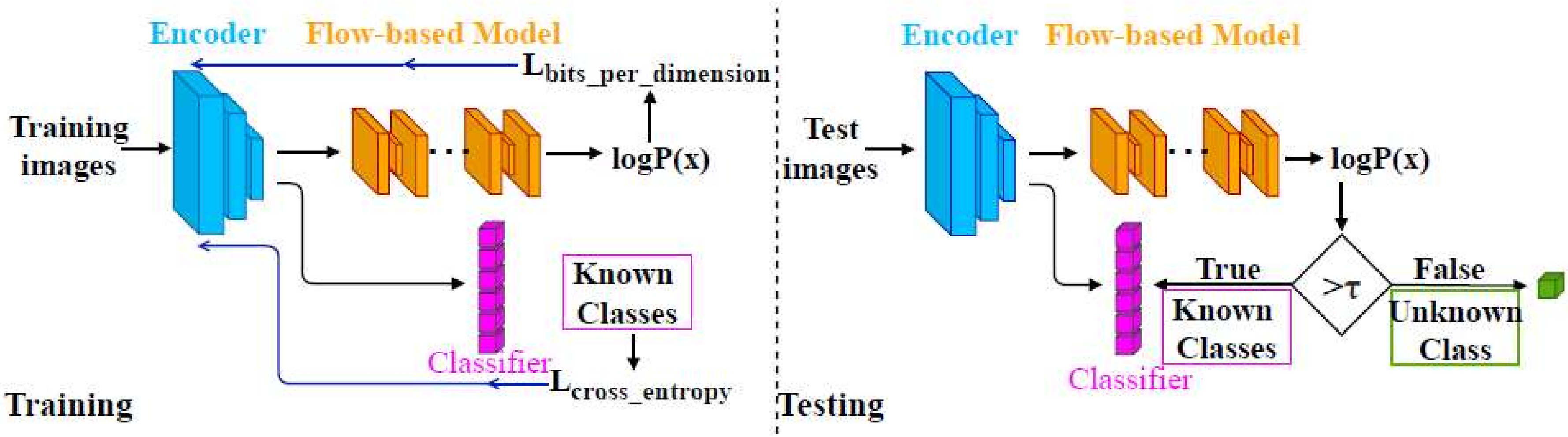
Approach 1: Distance-Based Methods

- Using k nearest neighbors in the z space does not work well in our experience

Approach 2: Density Estimation Methods

- Deep Density Estimation: Seeks to model $P(x)$ in the image space
  - These have not worked well
  - Le Lan & Dinh (2021) show that the representation is critical for density estimation. The image space lacks an appropriate "neighborhood structure"
- Methods that fit a classifier and then estimate a density $P(z)$
  - Open Hybrid (Zhang, Li, Guo, Guo, 2020)
  - Mahalanobis (Lee, Lee, Lee, Shin, 2018)

Open Hybrid: Classification + Density Estimation
(Zhang, Li, Guo, Guo, 2020)



- Residual Flow Deep Density Estimator
  - (Chen, Behrmann, Duvenaud, et al. NeurIPS 2019)
- Standard Cross-Entropy Supervised Loss
  - Claim: This helps focus $P(x)$ on relevant aspects of the images
- Anomaly Score: $A(x_q) = -\log P(x_q)$

Open Hybrid Results

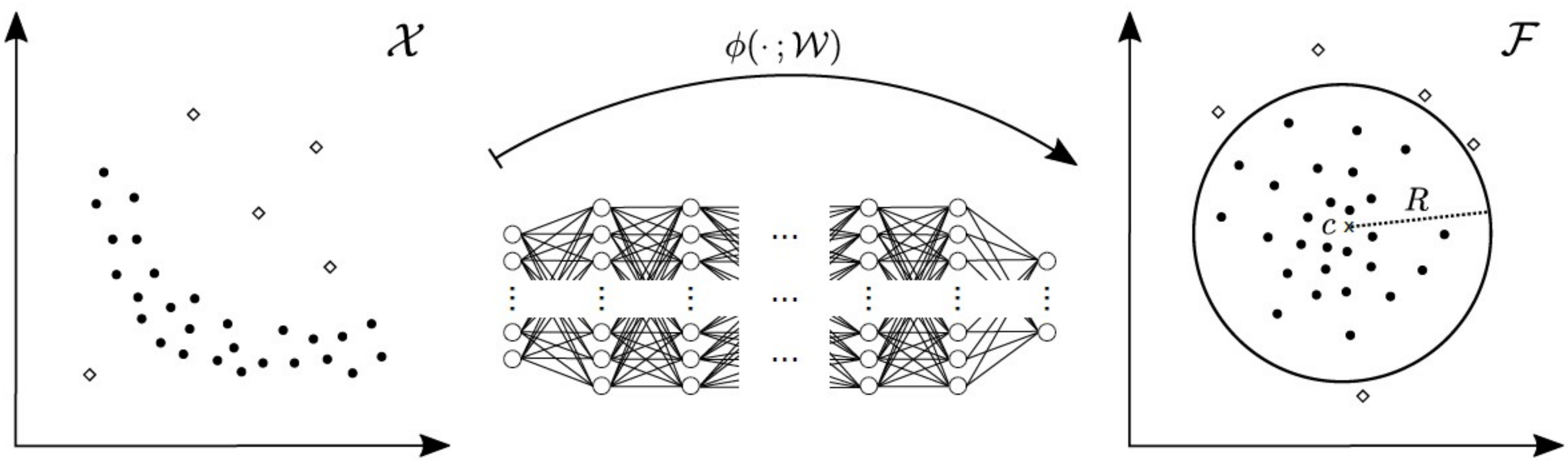
- 6 Known and 4 Unknown classes

AUC	MNIST	SVHN	CIFAR-10
$1 - \max_k \hat{p}(y = k \mid x_q)$	0.978	0.886	0.677
OpenHybrid: $-\log P(x_q)$	0.995	0.947	0.883

- 4 Known and many unknown classes drawn from CIFAR-100

AUC	CIFAR+10	CIFAR+50
$1 - \max_k \hat{p}(y = k \mid x_q)$	0.816	0.805
OpenHybrid: $-\log P(x_q)$	0.962	0.955

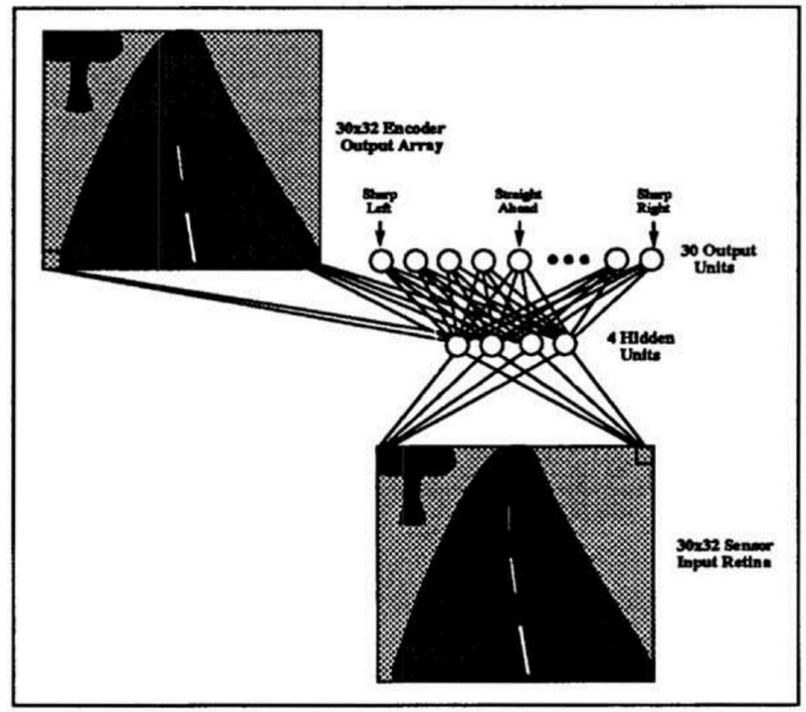
Approach 3: Density Quantile Methods: Deep SVDD
(Ruff, et al. ICML 2018)



- The method is somewhat tricky to work with
  - Set c as the mean of a small set of points passed through the untrained network
  - No bias weights
  - These help prevent "hypersphere collapse"

Approach 4: Reconstruction Methods

- NavLab self-driving van (Pomerleau, 1992)
  - Primary head: Predict steering angle from input image
  - Secondary head: Predict the input image ("auto-encoder")
  - $A(x_q) = \|x_q - \hat{x}_q\|$
  - If reconstruction is poor, this suggests that the steering angle should not be trusted
- Principle: Anomaly Detection through Failure
  - Define a task on which the learned system should fail for anomalies



Pomerleau, NIPS 1992

Deep Autoencoders

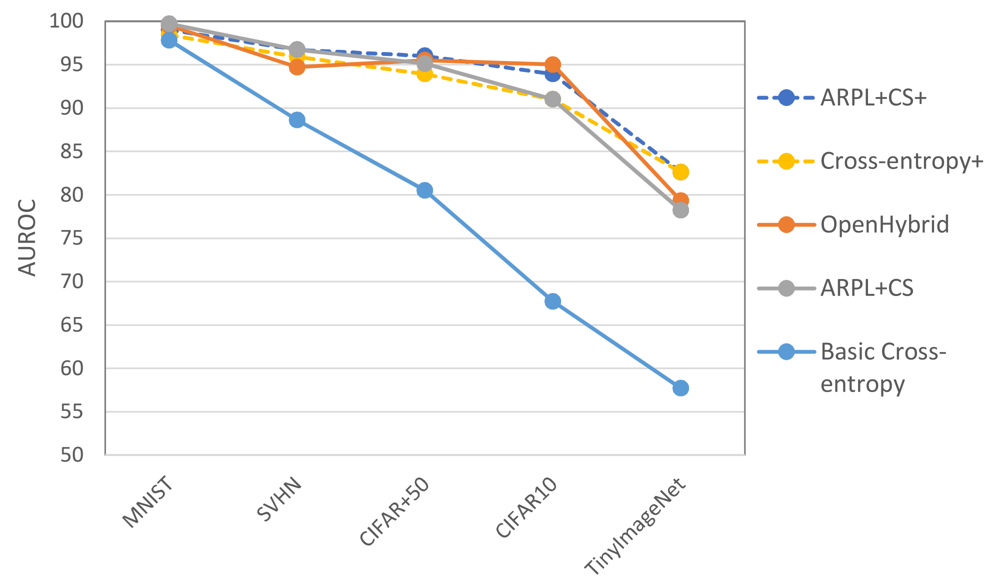
- These have generally not worked well for anomaly detection
- It is difficult to keep the autoencoder from learning to be a general image compression algorithm
    - It doesn't fail on novel images!

Outline

- Theoretical Approaches to Anomaly Detection
- Deep Anomaly Detection in Computer Vision
- Anomaly Detection based on Supervised Classifier Logit Scores

A Surprising Finding: Deep Classifiers Achieve Excellent Anomaly Detection

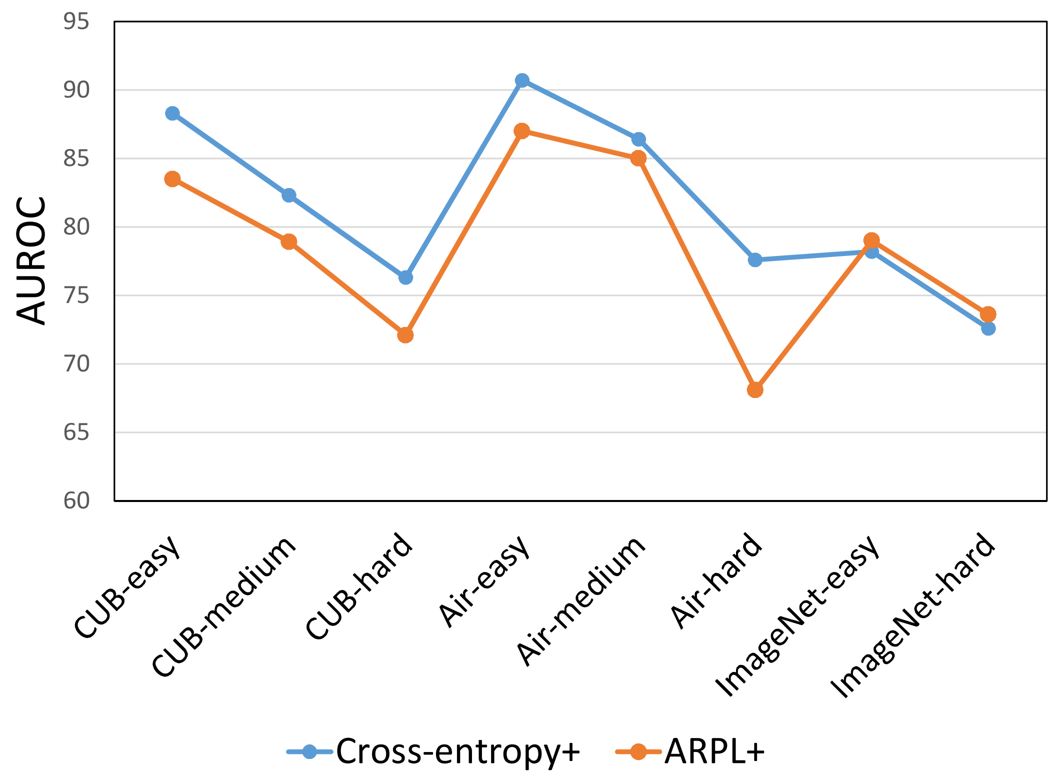
- Vaze, Han, Vedaldi, Zisserman (2020): "Open Set Recognition: A Good Classifier is All You Need"
  - arXiv 2110.06207
- Carefully train a classifier using the latest tricks
  - Standard cross-entropy combined with the following:
    - Cosine learning rate schedule
    - Learning rate warmup
    - RandAugment augmentations
    - Label Smoothing
- Anomaly score: max logit
  - $-\max_k \ell_k$



Protocol from Neal et al. (2018)

Vaze, et al.: Three Large Open Set Benchmarks

- Novel class difficulty based on semantic distance
  - CUB: Bird species
  - Air: Aircraft
  - ImageNet

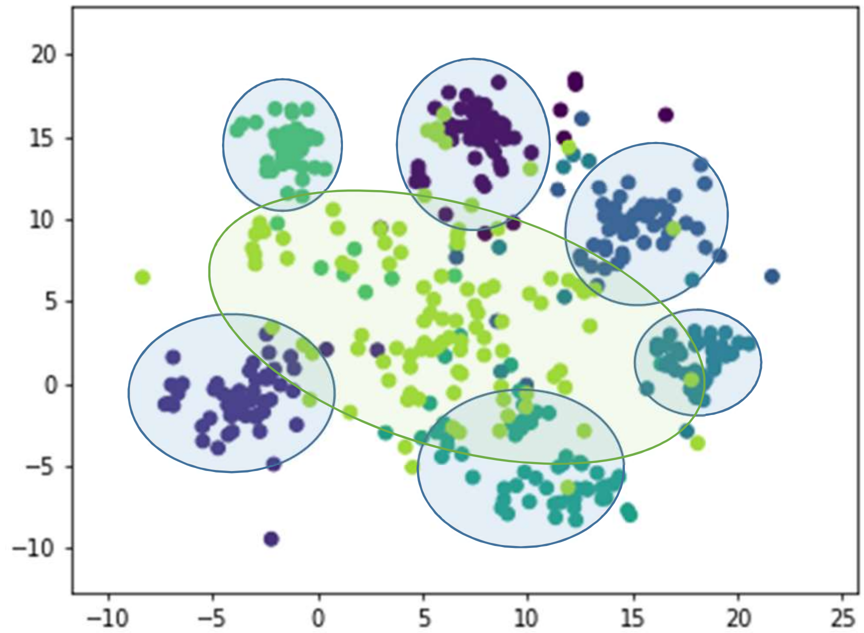


Why?

How are open set images represented by deep learning?

- DenseNet with 384-dimensional latent space.
- CIFAR-10: 6 known classes, 4 novel classes
- UMAP visualization
- Light green: novel classes
- Darker greens: known classes

- Note that many novel classes stay toward the center of the space; others overlap with known classes
- Training was not required to “pull them out” so that they could be discriminated

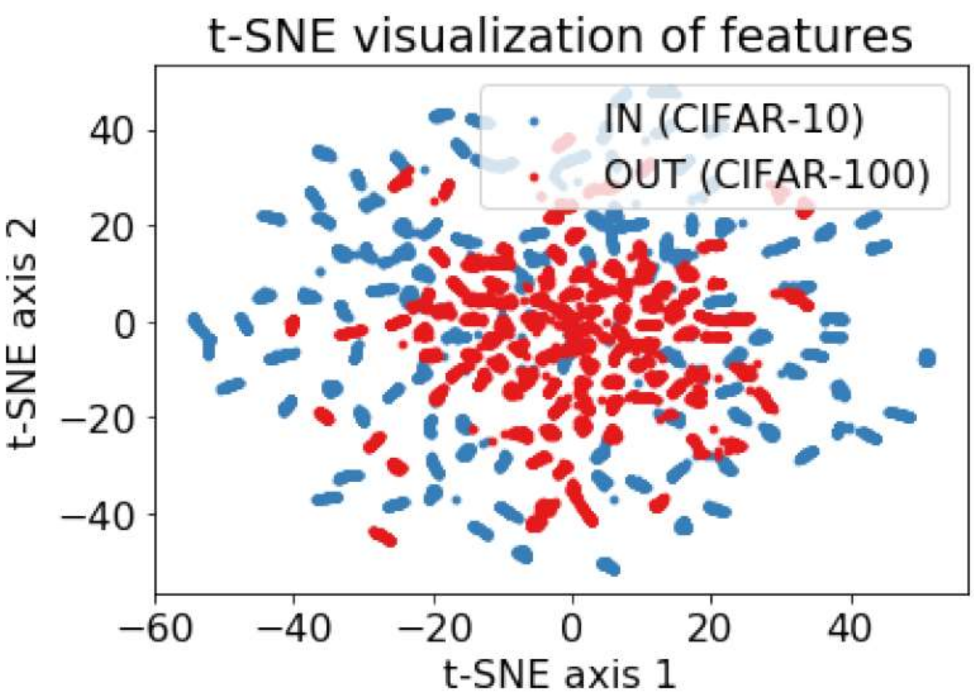


6 Known Classes

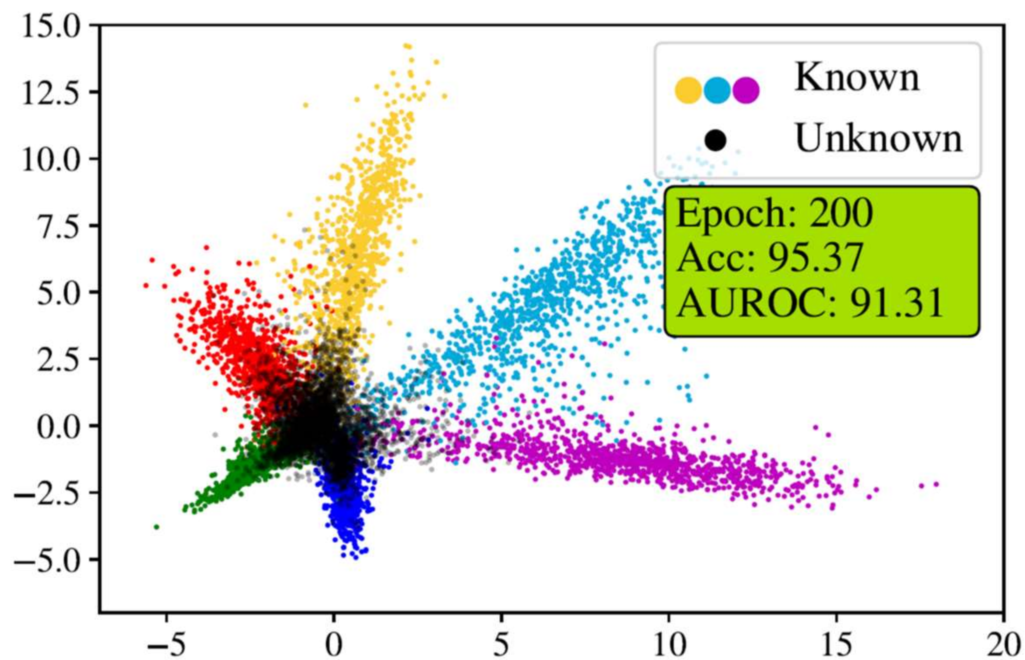
4 Novel Classes

Alex Guyer

Similar Results from Other Groups



[Tack, et al. NeurIPS 2020]



[Vaze, et al. arXiv 2110.06207]

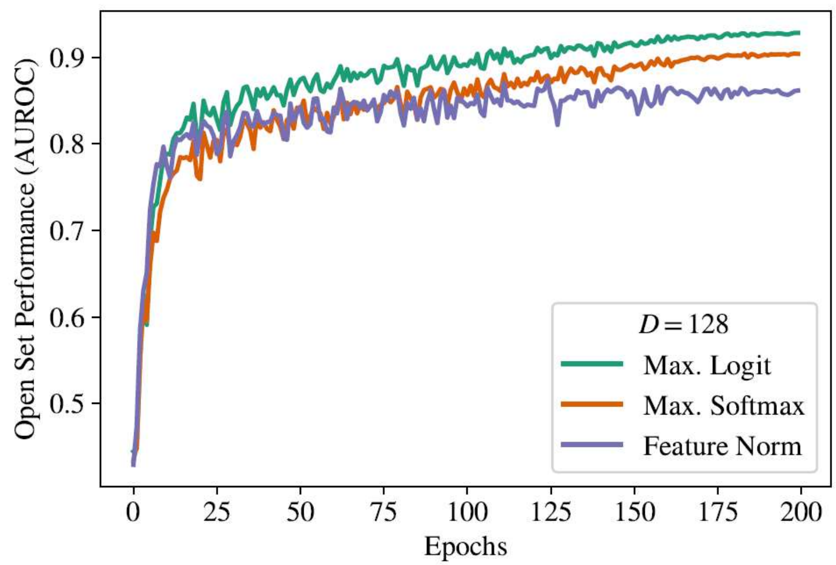
The Familiarity Hypothesis

The network doesn't detect novelty, it detects the absence of familiarity

- Convolutional neural network learns "features" that detect image patches relevant to the classification task
- The logit layer weights these features to make the classification decision
- Novel classes activate fewer of these features, so their activation vectors are smaller
- Hypothesis: The networks don't detect that an elephant is novel because of trunk and tusks but because its head doesn't activate known features

Initial Evidence

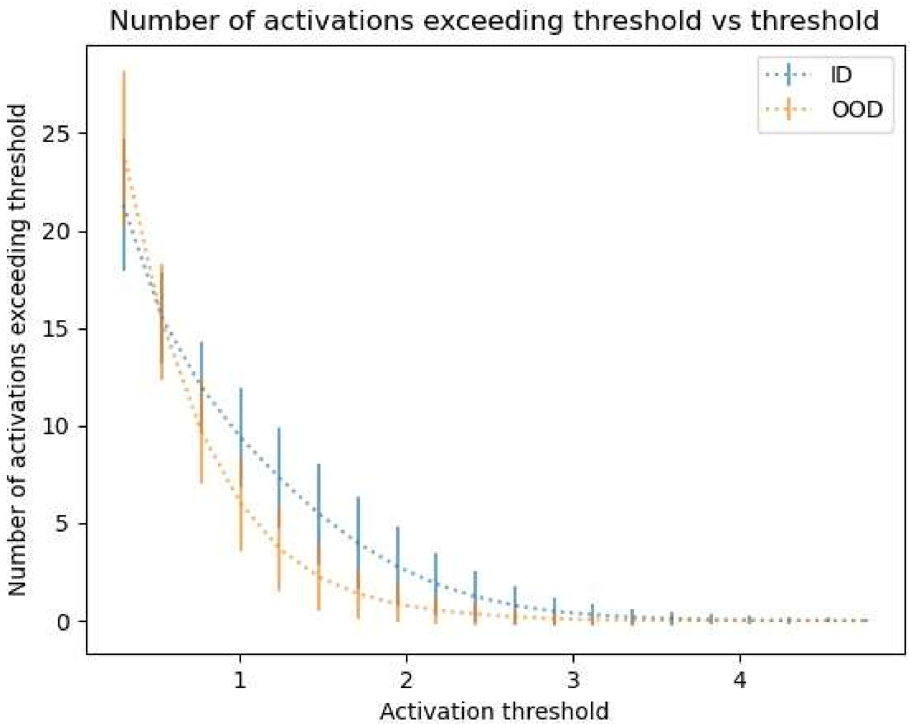
- Maximum logit is better than max softmax probability or the norm of z



Vaze, et al. 2021

Initial Evidence (2)

- CIFAR 10: 6 known classes; 4 novel classes
- DenseNet (z has 324 dimensions)
- Activation threshold θ
- Count number of features whose activation exceeds θ
- OOD images activate fewer features

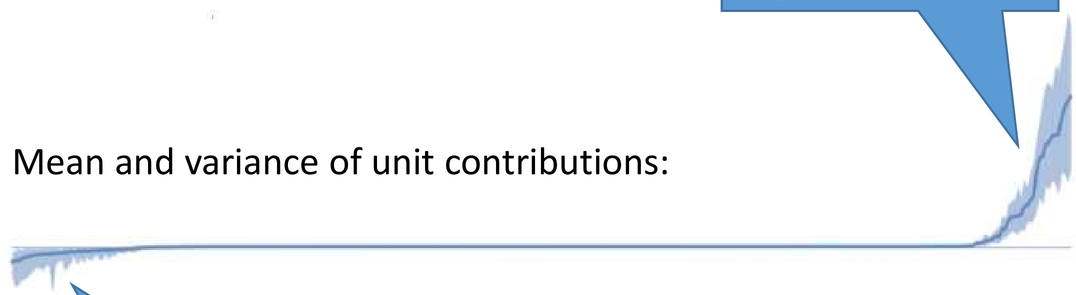


Alex Guyer (unpublished)

Max Sum of Positive Contributions

- Let w_k be the vector of weights for the logit of class k
$$\ell_k = w_k \cdot z$$
- Define the *contribution* of feature j to the logit for class k
$$c_{jk} = w_{jk} z_j$$
- Sort in ascending order
- Sum of positive contributions
$$s_k = \sum_j \max(0, c_{jk})$$
- Claim: $\max_k s_k$ will be an even better anomaly detector than max logit

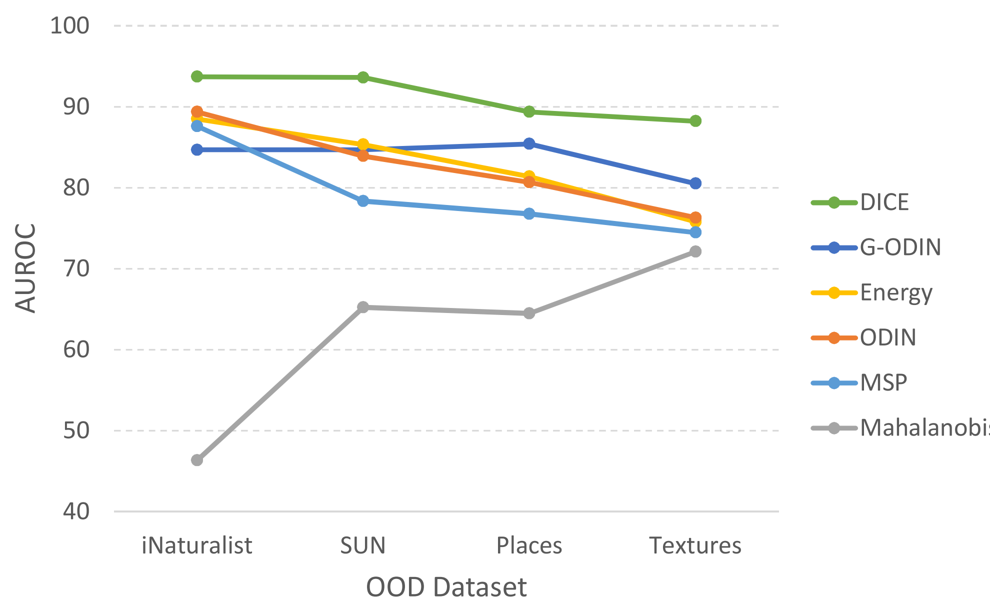
Mean and variance of unit contributions:



Sun & Li (2021) Did almost this experiment

- DICE anomaly score: sum of the top 10% of the contributions
  - MSP: $\max_k \hat{P}(y = k|x)$
  - Energy: $\log \sum_k \exp \ell_k$
  - ODIN and G-ODIN modify the Softmax

- Sun & Li (2021) “On the Effectiveness of Sparsification for Detecting the Deep Unknowns” arXiv 2111.09805



Can we expect computer vision systems to perceive things they have not been trained on?

- Colin Blakemore and Grahame F. Cooper. "Development of the brain depends on the visual environment." *Nature* (1970): 477-478.
  - Kittens raised in environments with only horizontal or only vertical lines
  - "They were virtually blind for contours perpendicular to the orientation they had experienced."

Source: Li Yang Ku
https://computervisionblog.wordpress.com/2013/06/01/cats-and-vision-is-vision-acquired-or-innate/

Possible Paths Forward

- Train on as many object categories as possible
  - Training on snakes might allow detection of trunks
- Systematically synthesize “natural parts”
  - Could we synthesize tusks, trunks if we had never seen them before?
  - Train on these to develop feature detectors for them
- Could a system detect “interesting image content” that was not activating learned features?

Summary

Classic Anomaly Detection Methods

- Distance
- Density estimation
- Density quantile estimation
- Reconstruction

Computer Vision Methods

- Well-trained Classifier works as well as deep anomaly detection methods
- Familiarity Hypothesis explains this and suggests improvements

Citations

- Bendale, A., & Boult, T. (2016). Towards Open Set Deep Networks. In CVPR 2016 (pp. 1563–1572). http://doi.org/10.1109/CVPR.2016.173
- Blakemore, C., & Cooper, G. F. (1970). Development of the brain depends on the visual environment. *Nature*, *228*(5270), 477–478. https://doi.org/10.1038/228477a0
- Breunig, M. M., Kriegel, H., Ng, R. T., & Sander, J. (2000). LOF: Identifying Density-Based Local Outliers. *ACM SIGMOD 2000 International Conference on Management of Data*, 1–12.
- Chen, R. T. Q., Behrmann, J., Duvenaud, D., & Jacobsen, J.-H. (2019). Residual Flows for Invertible Generative Modeling. *ArXiv*, *1906.02735*(v1), 1–18. http://arxiv.org/abs/1906.02735
- Chen, G., Peng, P., Wang, X., & Tian, Y. (2021). Adversarial Reciprocal Points Learning for Open Set Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–17. https://doi.org/10.1109/TPAMI.2021.3106743
- Emmott, A. F., Das, S., Dietterich, T., Fern, A., & Wong, W. (2013). Systematic Construction of Anomaly Detection Benchmarks from Real Data. KDD 2013 Workshop on Outlier Detection (ODD-2013), 6.
- Emmott, A. F. (2020). A Benchmarking Study of Unsupervised Anomaly Detection Algorithms. MS Thesis. School of EECS, Oregon State University.
- Guha, S., Mishra, N., Roy, G., & Schrijvers, O. (2016). Robust Random Cut Forest Based Anomaly Detection On Streams. *Proceedings of The 33rd International Conference on Machine Learning*, *48*. http://jmlr.org/proceedings/papers/v48/guha16.pdf
- Kim, J., & Scott, C. D. (2012). Robust Kernel Density Estimation. *Journal of Machine Learning Research*, *13*, 2529–2565.

Citations (2)

- Kriegel, H.-P., Schubert, M., & Zimek, A. (2008). Angle-based outlier detection in high-dimensional data. *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 444–452. https://doi.org/10.1145/1401890.1401946
- Larios, N., Soran, B., Shapiro, L., Martínez-Muños, G., Lin, J., Dietterich, T. G. (2010). Haar Random Forest Features and SVM Spatial Matching Kernel for Stonefly Species Identification*. IEEE International Conference on Pattern Recognition (ICPR-2010).*
- Lan, C. Le, & Dinh, L. (2020). Perfect density models cannot guarantee anomaly detection. *ArXiv*, *2012.03808*(v1), 1–15. http://arxiv.org/abs/2012.03808
- Lee, K., Lee, K., Lee, H., & Shin, J. (2018). A simple unified framework for detecting out-of-distribution samples and adversarial attacks. *Advances in Neural Information Processing Systems*, NeurIPS2018, 7167–7177. http://arXiv.org/abs/1807.03888
- Lin, J., Larios, N., Lytle, D., Moldenke, A., Paasch, R., Shapiro, L., Todorovic, S., Dietterich, T. (2011). Fine-Grained Recognition for Arthropod Field Surveys: Three Image Collections. *First Workshop on Fine-Grained Visual Categorization (CVPR-2011)*
- Liu, S., Garrepalli, R., Dietterich, T. G., Fern, A., & Hendrycks, D. (2018). Open Category Detection with PAC Guarantees. *Proceedings of the 35th International Conference on Machine Learning, PMLR*, *80*, 3169–3178. http://proceedings.mlr.press/v80/liu18e.html
- Liu, S., Garrepalli, R., Hendrycks, D., Fern, A., & Hendrycks, D., Dietterich, T. G. (to appear) PAC Guarantees and Effective Algorithms for Detecting Novel Categories. *Journal of Machine Learning Research.*
- Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. *2008 Eighth IEEE International Conference on Data Mining*, 413–422. https://doi.org/10.1109/ICDM.2008.17
- Lytle, D. A., Martínez-Muñoz, G., Zhang, W., Larios, N., Shapiro, L., Paasch, R., Moldenke, A., Mortensen, E. A., Todorovic, S., Dietterich, T. G. (2010). Automated processing and identification of benthic invertebrate samples. *Journal of the North American Benthological Society, 29*(3), 867-874.

Citations (3)

- Pevný, T. (2015). Loda: Lightweight on-line detector of anomalies. *Machine Learning*, *November 2014*. https://doi.org/10.1007/s10994-015-5521-0
- Pomerleau, D. A. (1993). Input Reconstruction Reliability Estimation. *Proceedings of NIPS 1993*, 279–286.
- Rudd, E. M., Jain, L. P., Scheirer, W. J., & Boult, T. E. (2017). The Extreme Value Machine. *ArXiv*, *1506.06112*, 1–12. https://doi.org/10.1109/TPAMI.2017.2707495
- Ruff, L., Vandermeulen, R. A., Shoaib, D., Binder, A., Emmanuel, M., & Kloft, M. (2018). Deep One-Class Classification. *International Conference on Machine Learning (ICML 2018)*, 10.
- Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Computation*, *13*(7), 1443–1471. https://doi.org/10.1162/089976601750264965
- Sun, Y., & Li, Y. (2021). On the Effectiveness of Sparsification for Detecting the Deep Unknowns. *ArXiv*, *2111.09805*(v1). http://arxiv.org/abs/2111.09805
- Tack, J., Mo, S., Jeong, J., & Shin, J. (2020). CSI: Novelty Detection via Contrastive Learning on Distributionally Shifted Instances. *Advances in Neural Information Processing Systems (NeurIPS 2020)*.
- Tax, D., & Duin, R. (2004). Support vector data description. *Machine Learning*, 45–66. http://link.springer.com/article/10.1023/B:MACH.0000008084.60811.49
- Vaze, S., Han, K., Vedaldi, A., & Zisserman, A. (2021). Open-Set Recognition: A Good Closed-Set Classifier is All You Need. *ArXiv*, *2110.06207*(v1), 1–23. http://arxiv.org/abs/2110.06207
- Wagstaff, K. L., Lanza, N. L., Thompson, D. R., Dietterich, T. G., & Gilmore, M. S. (2013). Guiding Scientific Discovery with Explanations using DEMUD. *AAAI 2013*.
- Zhang, H., Li, A., Guo, J., & Guo, Y. (2020). Hybrid Models for Open Set Recognition. *ArXiv*, *2003.12506*(v1), 1–17. http://arxiv.org/abs/2003.12506