Contrastive-GMMs for Open-World Learning

Stephen Scarano Preston Yee Kobe Falus

April 19, 2023

Project description. The success of deep learning architectures has attracted the attention of mainstream culture, a testament to its success and impact. However, conventional methodologies maintain that success under strict conditions—ones not always feasible in a dynamic, open environment.

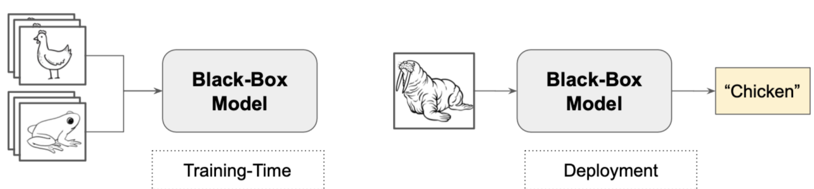
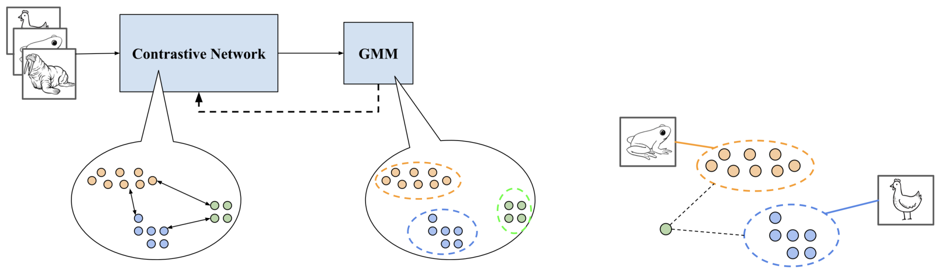


Figure 1: An open-world deployment risks encounters with classes not seen during training time.

The most successful deep learning projects have traditionally been those tested in a strictly closed-setting; that is, the model is evaluated with respect to some number of classes observed during training. This is not always tractable. Even recent ambitious breakthroughs in deep learning–namely LLMs–are prone to hallucination and a lack of interpretability [2]. Our approach seeks to leverage empirical methods of model selection and traditional machine learning to ease these sharp edges–while also opening the model setting to new potential classes.

As shown in Figure 2a, our network is composed of a Contrastive block and a GMM block. The former identifies features to "contrast" against instances of other classes that are simultaneously similar to peers of their own class. The GMM block determines how easy it is to draw a simple decision boundary around these points, while also instantiating a new "novel" label to identify instances sufficiently far from any class Gaussian (see Figure 2b).



(a) Proposed Architecture (b) Novelty Detection

Figure 2: Three simple graphs

The major speculation of this approach is whether or not features of the known labels will generalize to those of the unknown ones. It is for this reason our model tracks outliers dynamically during test-time: after GMM classification, we leverage AIC criterion to determine whether or not a model with a greater number of components is better suited to the accumulated novel instances. If so, we retrain the Contrastive block using the clustered novel points as a functional "new class".

Related work. We plan to implement the above architecture using a Multi-View CNN framework, though the choice of network is somewhat arbitrary. On the last page of this document, we list the literature we plan to review, broken down by area of relevance.

Datasets and evaluation. We believe the *ShapeNet* repository [Link], with its numerous class instances of varying similarity (i.e, chairs and stools are more similar than planes and people), is an ideal dataset to test. Evaluation is a uniquely complex problem in novelty and open-world learning work, and we expect to discover greater nuance and methods as we review the literature; however, for now we will present preliminary criteria.

We will evaluate our model with respect to two tasks, one a function of the other:

1. **Detection of novel instances:** We follow the formulation of Open Set Recognition (OSR) from [10] in which we evaluate with respect to Known-Known Classes (KKC) and Unknown Unknown Classes (UUC). Our model has observed instances of the former classes, but not the

latter. Since UUC instances may outnumber KKCs, a trivial modification to standard accuracy,

$$ACC = \frac{(TP_{KKC} + TN_{KKC}) + (TP_{UUC} + TN_{UUC})}{(FP_{KKC} + FN_{KKC}) + (FP_{UUC} + FN_{UUC})}$$

may inform us very little of the model's ability to predict KKC instances: we may have simply built a "novelty detector" in the most literal sense. We instead utilize a normalized-accuracy criterion suggested by the authors of [11]:

$$NA = \lambda_r(ACC_{KKC}) + (1 - \lambda_r)(ACC_{UUC})$$

Where r is a regularization parameter that weighs the importance of accuracy for the KKC and UUC instances. As argued by [69] our F1 measure can be largely left the same, where we only consider KKC classes except for False Negatives (FN) and False Positives (FP).

2. **Open World Learning:** we build upon the metrics provided in (i), by considering UUC instances not as a class, but rather as a set of classes.

Split of the work. Since the project is relatively ambitious (and we'd like to evaluate the method cohesively), we think a three-person team is reasonable. The blueprint thus far is:

- Stephen Scarano: architecture development, writing
- Preston Yee: evaluation, writing
- Kobi Falus: literature review, writing, evaluation

All writers will contribute towards the presentation and writing of the final report.

0.1 Literature

1 "Multi-view Convolutional Neural Networks for 3D Shape Recognition" [Link]

2 "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead" [Link]

3 "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" [Link]

And integrate recent work in the subfield of contrastive learning:

4 "Supervised Contrastive Learning" [Link]

5 "Big Self-Supervised Models are Strong Semi-Supervised Learners" [Link]

3 [6] "A Simple Framework for Contrastive Learning of Visual Representations" [Link]

And obviously in the domain of Open-World Learning:

7 "Open-world Machine Learning: Applications, Challenges, and Opportunities" [Link]

8 "A Critical Evaluation of Open-World Machine Learning" [Link]

9 "A Review of Open-World Learning and Steps Toward Open-World Learning Without Labels" [Link]

10 "Recent Advances in Open Set Recognition: A Survey" [Link]

11 "Nearest neighbors distance ratio open-set classifier" [Link]

And build upon the foundations of existing empirical and statistical methods:

12 "The expectation-maximization algorithm" [Link]

13 "The Mahalanobis distance" [Link]