

How to Know What You Don't Know: Open Set Deep Learning

Jordan Spell
jdsPELL@iu.edu

Josiah Keime
keimej@iu.edu

Kyle Firestone
kfiresto@iu.edu

Abstract—In this project we aim to explore image classification with open set deep learning networks, that is to say networks which are presented with image classes not contained in the training set and are tasked with being able to identify that these are unknown classes as opposed to confidently predicting these images belong to classes in the training set. We will further explore if these models cluster distinct unknown classes sufficiently different from each other that they can perform zero-shot or few-shot learning on the unknown classes.

I. INTRODUCTION

Deep learning networks have immensely improved image classification over the past decade with numerous models achieving over 90% top-1 accuracy on the ImageNet data set [1]. However, for many of these models they lack generalization and can struggle to perform in real-world scenarios where there are certainly more classes than the model was exposed to in its training set. For example, in autonomous driving it is unlikely that we could label every class a vehicle is likely to encounter. So, how then, does a model behave when exposed to an image class that was not in the original data set? In some cases the model will incorrectly make a confident prediction of a class that was in the training set [2]. Even if there is no confident prediction, unless a prediction threshold rule has been put in place for the model, an incorrect class will still be predicted. This is the problem with closed set classification, the model assumes everything is a member of the training set classes. Open set models on the other hand are built with the knowledge that a model will see classes not in the training set and will train it to recognize that it does not know what class an image is.

Open set models were first considered outside of deep learning just as the deep learning revolution was taking place [3]. This approach utilized an SVM to determine how far was to far from a classification boundary and considered those as unknown classes. Since then several deep learning models have been developed looking at open set learning [4, 5]. The aim of this project is to explore existing open set deep learning frameworks and investigate improving these models. Specifically we will look at high dimensional embeddings of features learned from a network and use clustering in this embedded space to determine if an example is an unknown class. Furthermore, we aim to see if these models are capable of performing zero-shot or few-shot learning on these unknowns.

II. PROJECT GOALS

- Recreate an open set image classification model from the literature.
- Iterate on model with the goal of maximizing separation of features in the embedding space to best be able to succeed in zero-shot or few-shot learning.
- Test robustness of models on adversarial data.

III. TEAM MEMBER RESPONSIBILITY

Responsibilities are subject to change and will be very fluid with frequent collaboration between team members on tasks.

Team Member	Task
All Members	Recreate open set model from literature
Jordan	Work on feature embeddings Zero-shot learning
Josiah	Add openMax to existing classifier Iterate on models Test with adversarial data
Kyle	Data set creation Iterate on model Few-shot learning

IV. DATA SET

We plan to begin by using ImageNet with a subset of classes removed to serve as the unseen classes for our models. If this proves to be problematic or we are getting very good performance we will examine other image data sets.

REFERENCES

- [1] Papers with Code - ImageNet Benchmark (Image Classification) — paperswithcode.com.
- [2] A. Nguyen, J. Yosinski, and J. Clune. (2015) Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on. IEEE.
- [3] Scheirer, W., Rezende Rocha, A., Sappkota, A., & Boulton, T. (2013). Toward Open Set Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(7), 1757-1772.
- [4] A. Bendale, & T. E. Boulton (2016). Towards Open Set Deep Networks. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 1563-1572). IEEE Computer Society.
- [5] Shu, Y., Shi, Y., Wang, Y., Huang, T., & Tain, Y. (2020). "P-ODN: Prototype-Based Open Deep Network for Open Set Recognition." Scientific Reports, vol. 10, no. 1, p. 7146.