Demystify the Unknowns: Unknown types, Methods and Limitations

Sarthak Jain
Los Gatos High School
jainsarthak101@gmail.com

Shu Kong
Carnegie Mellon University
shuk@andrew.cmu.edu

Abstract

Computer Vision models will often face image data that is out-of-distribution from its train set. Coined as Open-Set Recognition, previous research has created algorithms that exploit K-way classification to classify such out-of-distribution images. 1) MaxLogits utilizes the values obtained from the final nodes of trained CNN classifier while 2) OpenGAN generates its own fake data to train a discriminator that can focus on the features of an image. In our paper, we train OpenGAN and MaxLogits when K=180 and when K=6. We then study the performance of these four models on 3 different types of out-of-distribution data. We find that OpenGAN outperforms MaxLogits in the Gaussian/Uniform Noise and Noise as Unknown Evaluations although the difference in AUPR and AUROC is wider when K=180. For Cross-Dataset Evaluation, OpenGAN again outperforms MaxLogits when the out-of-distribution set is SVHN or EuroSAT, but when it is MNIST, MaxLogits performs better.

1. Introduction

The training set for a neural network will not represent the entire world and possibility of test-time data. Such test-time data that has significant differences between itself and the training set as a whole can be categorized into a never-before-seen category. Different architectures have been designed to determine if test data falls in the category of in-distribution or out-of-distribution data. This field is known as Unknown Detection or Open-Set Recognition.

Binary Classifiers have been implemented to address such challenges. These classifiers are typically trained on an outlier set with classes that are disjoint to the in distribution set. The outlier data set is labeled differently from the in-distribution data set and a simple one output node network is trained to better classify new images as out-of-distribution or in-distribution. However, one issue with this approach is that the binary classifier has a large tendency to overfit to the out-of-distribution set and therefore struggles when faced with test-data that has not been seen in both the out-of-distribution and the in-distribution set [2].

Another approach, called MaxLogits or MaxSoftmax, entails training a K-way classifier network on the in-distribution data set. The captured output of a K-way classification network, whether it is K raw logits or K softmax probabilities, can be used as a good indicator for Open-Set Recognition. Importantly, previous research has shown that removing the softmax and just extracting the logits of the neural network gets more accurate results. [6]

Lastly, one of the latest approaches known as OpenGAN [3] utilizes a Generative Adversarial Neural Network [1] to train a discriminator that can delineate whether the features of an image are in distribution or out of distribution. OpenGAN trains on features of images that are created through passing that image into a K-way classification model that is exactly same as the one which could be used in a MaxLogits or MaxSoftmax case. Unlike other architectures, OpenGAN both generates its own out-of-distribution features that approaches the likelihood of the in-distribution features and also leverages an outlier dataset.

Out-of-distribution data can appear in many formats such as high level noise that is semantically different from the in-distribution data or low-level noise that is semantically similar to the in-distribution data but still contains a different feature composition. Additionally the architecture that is better suited for Open-Set Recognition depends heavily on the feature composition and number of classes in the in-distribution set. We attempt to demystify architecture performance when testing on different types of out of distribution data while trained on certain in-distribution sets. The two architectures we focus on are Max-Logits and OpenGAN since they have produced some of the best results in the field.

2. Methodology

We focus on two recent state-of-the-art methods below. **OpenGAN** generates its own out-of-distribution set that approaches the likeness of the in-distribition set. OpenGAN also utilizes an auxiliary outlier set for training. Unlike traditional GANs, however, OpenGAN trains on the features collected from a K-way classification and generates features. The K-way classification network has already been trained on the in-distribution set. Then the feature generator is trained in tandem with the feature discriminator with in-distribution image features being given the label of 1 while out-of-distribution image features and generated features are given the label of 0. **MaxLogits** is a different application of the K-way classification network. It takes the raw logits of the output K nodes of a K-way classifier and uses those values as an indicator of which an image is in-distribution or out-of-distribution.

3. Experiment

3.1. Protocols

We evaluated OpenGAN and Max Logits when trained on datasets with high K classes and low K classes. The high K classes dataset we used was tinyImageNet [4]. The train set we used consisted of the first 180 classes of tinyImageNet. The low K classes dataset we used was CIFAR. We used the first 6 classes of CIFAR.

The Out of Distribution images were selected to be disjoint from ImageNet and as a result disjoint from tinyImageNet and CIFAR which are categorical subsets to ImageNet. Below is a list of every single out of distribution dataset that was used for evaluation along with an explanation to justify its inclusion.

- **Cross-Dataset Evaluation**: For the semantically different evaluation datasets, we used test SVHN [5] and a three channel concatenated test MNIST. Both Datasets measure how well a model would distinguish a data set semantically different than the train-set as out of distribution.

- **Gaussian or Uniform Noise Evaluation**: For the noise datasets, we used images entirely made of pixels randomly sampled from Gaussian and Uniform noise distribution. This created dataset served to test a model's accuracy when put up against images with no features at all. Since both Max Logits and OpenGAN utilize a K-way classification backbone to extract features, such an evaluation would test if these algorithms efficacy dependence on the presence of such features.

- **Noise as Unknown**: For feature containing data sets with noise, we took images from the test version of the in-distribution set (tinyImageNet) and added a pair of strong or weak transforms to those images. The light transform consisted of Enhance Contrast and Gaussian Blur while the dark transform consisted of Flipping, Enhance Contrast, a stronger Gaussian Blur, and Rain transform. Its inclusion is to test how well models can distinguish if an image might have a lot of noise despite being the same class as an image from the train set. Models that perform well on this dataset serve to be well used when images taken from a system like an autonomous vehicle may not contain any unknown objects but are still significantly different to the images and feature composition of the train-set.

3.2. Metrics

The two metrics we report are Area Under Precision Recall Curve (AUPR) and Area Under Receiver Operating Curve (AUROC). Both these metrics are ranking based evaluation systems that have been used by previous Open-Set Recognition work. **AUPR** is the probability that a point listed above a positive classified point is also positive. **AUROC** measures the performance of a binary classifier across varying confidence thresholds. The test in-distribution dataset for both OpenGAN and MaxLogits models consisted of an unseen version of the initial in-distribution (so either test CIFAR or test tinyImageNet depending on what the OpenGAN or MaxLogits model was trained on). Depending on the protocol, the test out-of-distribution model changes (See 3.1).

3.3. Implementations

Every K-way Model and OpenGAN model was trained using only one GPU. **The K-way model for tinyImageNet** was trained with 50 epochs. For training this K-way model we utilize Stochastic Gradient Descent with a learning rate of 0.003, weight decay of 0.0005, and momentum of 0.9. We also utilize a Lambda learning rate scheduler with the lambda following cosine. Since we have 180 classes taken from out tinyImageNet dataset, this K-way model has 180 output nodes. The augmentations that were used for the training data were random cropping and random horizontal flipping.

The OpenGAN model for tinyImageNet used the aforementioned trained K-way model to collect features from an input image. Then the Discriminator and Generator of the OpenGAN were trained for 20 epochs. For training the OpenGAN, we had two optimizers: one for the Generator and one for the Discriminator. The Generator optimizer has a base learning rate of 0.0001 while the Discriminator's learning rate is a 0.0001/1.5. The OpenGAN was trained on both the testTinyImagNet and the SVHN dataset (train out-of-distribution dataset). The testTinyImageNet was given the same transforms as when it was used for training the K-way classifier. The SVHN dataset was randomly horizontally flipped and resized.

3.4. Results

When using a high class dataset (K = 180) for training. OpenGAN usually outperforms Max-Logits as an open-set recognizer. For Cross-Dataset Evaluation OpenGAN has better AUPR and AUROC for the test cases of SVHN vs testTinyImageNet and EuroSAT vs tinyImageNet. However, for MNIST (where grey scale channels are cloned into 3 channels) vs Tiny, Logits has better AUPR and AUROC. However, the MaxLogits performance is very close to a random 50/50 guess indicating both our trained OpenGAN and MaxLogits are not suited for this classification task. For the Noise evaluations, OpenGAN performed significantly better than Max-Logits when tested with Gaussian Noise and Uniform Noise. Finally for the Noise as Unknown, OpenGAN again outperforms Max Logits in both heavy transforms and light transforms.

Table 1. The in-distribution dataset was testTinyImageNet. K-way model pretrained on ImageNet and Finetuned on trainTinyImageNet. The outlier-set was trainSVHN (AUROC / AUPR).

open-set	MaxLogits	OpenGAN
SVHN	26.19 / 40.12	**99.97 / 99.97**
EuroSAT	82.18 / 84.40	**94.85 / 94.86**
MNIST	**54.46 / 55.86**	22.20 / 38.43
Normal Noise	20.57 / 44.13	**100.0 / 100.0**
Uniform Noise	26.90 / 46.20	**77.89 / 84.02**
Transforms 1 (heavy)	64.56 / 67.72	**87.37 / 83.20**
Transforms 2 (light)	57.73 / 57.72	**68.89 / 62.10**

When using a low class dataset (K = 6) for training, OpenGAN still generally outperforms Max-Logits. For Cross-Dataset Evaluation, OpenGAN again had a larger AUROC and AUPR than Max-Logits for SVHN and EuroSAT evaluation. However, interestingly for MNIST Max-Logits massively outperforms OpenGAN which is similiar to our results when K = 180. This suggests a motivation of creating an ensemble between Max-Logits and OpenGAN. Also similiar to the high class case, OpenGAN has a better AUROC and AUPR than MaxLogits for both the noise evaluations. For the Noise as Unknown Evaluations, OpenGAN again outperforms Max-Logits for both heavy and light transforms.

Table 2. The in-distribution dataset was testCIFAR. K-way model pretrained on ImageNet and Finetuned on trainCIFAR The outlier-set was trainSVHN (AUROC / AUPR).

open-set	MaxLogits	OpenGAN
SVHN	67.07 / 74.91	**89.18 / 88.37**
EuroSAT	77.60 / 84.05	**93.27 / 94.91**
MNIST	**93.41 / 94.43**	35.74 / 47.55
Normal Noise	80.90 / 85.75	**98.78 / 98.93**
Uniform Noise	71.16 / 81.79	**94.6 / 96.93**
Transforms 1 (heavy)	65.48 / 69.46	**75.62 / 71.77**
Transforms 2 (light)	56.31 / **56.95**	**57.78** / 54.60

4. Discussion and Future Work

Through various evaluations, we exemplify that OpenGAN generally outperforms MaxLogits on both Gaussian and Uniform Noise Evaluations and Noise as Unknown Evaluations irrespective of the whether it was trained on a high class (K = 180) or low class (K = 6) dataset. However, the benefit you get from OpenGAN is increasingly pronounced when using a high class dataset; therefore, suggesting that OpenGAN benefits from higher classes compared to MaxLogits. For Cross-Dataset Evaluation, The results were a little more mixed. OpenGAN outperformed MaxLogits on SVHN (train version was also the outlier set) and on EuroSAT; however, it did not outperform MaxLogits on MNIST for both K=180 and K=6. Therefore, this suggests that an ensemble network that combines the insight of MaxLogits and OpenGAN would be a more ideal solution to perform better in Cross-Dataset Evaluation tasks and area of future work.

One possible limitation is that we did not follow the standard open-set recognition literature that does not pretrain models, instead, we use pretrained models finetuned for closed-set recognition (although we guarantee that pretraining-set and the closed training set have disjoint classes). In the future, we plan to compare the results with and without pretraining.

References

[1] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020. 1

[2] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. *arXiv preprint arXiv:1812.04606*, 2018. 1

[3] Shu Kong and Deva Ramanan. Opengan: Open-set recognition via open data generation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 813–822, 2021. 1

[4] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015. 2

[5] Zhibin Liao and Gustavo Carneiro. Competitive multi-scale convolution. *arXiv preprint arXiv:1511.05635*, 2015. 2

[6] Sagar Vaze, Kai Han, Andrea Vedaldi, and Andrew Zisserman. Open-set recognition: A good closed-set classifier is all you need. *arXiv preprint arXiv:2110.06207*, 2021. 1