Out-of-Distribution Detection

DAKANNAGARI ROHITH REDDY
COMPUTER SCIENCE AND ENGINEERING
LOVELY PROFESSIONAL UNIVERSITY
12002425

Abstract

This paper introduces a novel approach to enhance the reliability of computer vision systems under previously unseen imaging conditions. The proposed method involves adapting camera parameters using a normalizing flow-based out-of-distribution detector. A small-scale study has been conducted, demonstrating that adjusting camera parameters based on this out-of-distribution detector results in an average improvement of 3 to 4 percentage points in key performance metrics, including mean Average Precision (mAP), mean Average Recall (mAR), and F1 score, for a YOLOv4 object detection model.

Additionally, this research showcases the feasibility of training a normalizing flow model for out-of-distribution detection using the COCO dataset, which is notably larger and more diverse than many existing benchmarks for evaluating out-of-distribution detection methods.

1.Introduction

In the realm of modern machine learning models deployed in real-world scenarios, a common challenge is dealing with out-of-distribution (OOD) inputs. These are samples that come from a distribution different from what the model has seen during its training, and they should not be predicted during testing. An ideal classifier not only correctly classifies known in-distribution (ID) samples but also has the capability to flag OOD inputs as "unknown." This underscores the significance of OOD detection, which is essential for determining whether an input belongs to the known distribution (ID) or falls outside of it (OOD), allowing the model to take appropriate actions.

Recent advancements in OOD detection have seen the development of various algorithms, including distance-based methods like those demonstrated by Lee et al. (2018), Tack et al. (2020), and Sehwag et al. (2021). These methods rely on feature embeddings extracted from models and typically assume that OOD samples are significantly distant from the in-distribution data. For example, Lee et al. modeled the feature embedding space using a mixture of multivariate Gaussian distributions and employed the maximum Mahalanobis distance for OOD detection, but these approaches make strong distributional assumptions about the underlying feature space.

This leads to the question: Can we utilize a non-parametric nearest neighbor approach for OOD detection? Unlike previous methods, the non-parametric approach doesn't make distributional assumptions about the feature space, providing more flexibility and generality. However, despite its simplicity, the non-parametric nearest neighbor approach has received limited attention. Despite a growing body of research in OOD detection, there has been no prior work demonstrating the effectiveness of a non-parametric nearest neighbor approach. This suggests that making this seemingly simple idea work is challenging.

In this paper, we challenge the conventional wisdom by presenting the first comprehensive exploration of and demonstration of the effectiveness of the non-parametric nearest-neighbor distance for OOD detection. To identify OOD samples, we compute the k-th nearest neighbor (KNN) distance between the embedding of a test input and the embeddings of the training set. We use a threshold-based criterion to determine if the input is OOD or not. In essence, we perform non-parametric level set estimation, dividing the data into two sets (ID vs. OOD) based on the deep k-nearest neighbor distance. KNN offers several advantages, including being distributional assumption-free, OOD-agnostic, easy to use, and model-agnostic.

Our exploration demonstrates both empirical effectiveness and theoretical justification. We show that a compact and normalized feature space is crucial for the success of the nearest neighbor approach for OOD detection. Extensive experiments indicate that KNN outperforms parametric approaches and scales effectively for large-scale datasets. Modern approximate nearest neighbor search implementations make this approach feasible even with vast databases. On a challenging ImageNet OOD detection benchmark, our KNN-based approach achieves superior performance while maintaining similar inference speeds to baseline methods. The simplicity and effectiveness of KNN make it a promising choice for real-world applications.

In summary, our contributions are as follows:

1. We introduce the first comprehensive study of non-parametric density estimation with nearest neighbors for OOD detection, highlighting its promise as a flexible and underexplored approach in the literature.

2. We demonstrate the superior performance of the KNN-based method across various OOD detection benchmarks, model architectures, and training losses, substantially reducing false positive rates compared to strong baseline methods.

3. We provide valuable insights into the key factors that make KNN effective in practice, including feature normalization and a compact representation space, supported by extensive ablations and experiments.

4. We offer theoretical analysis, demonstrating that KNN-based OOD detection can reject inputs equivalent to the Bayes optimal estimator, linking our method to the feature space and complementing our experiments in considering the universality of OOD data.

2. Related Work

Previous research has explored the applications of out-of-distribution (OOD) detection in autonomous systems. For instance, Wellhausen et al. [17] conducted anomaly detection on image data captured by a robot navigating real-world terrain. However, their primary focus was the evaluation of various OOD detection algorithms, rather than the practical implementation of OOD detection in real-world operational scenarios.

yuhas et al. [20] assessed OOD detection as an emergency braking system for an autonomous car, although their evaluations were not carried out in genuine real-world operational settings but rather on a custom test track.

McAllister et al. [12] conducted real-world crash avoidance experiments with an autonomous vehicle. However, their approach did not directly involve OOD detection. Instead, they utilized a variational autoencoder to generate in-distribution samples from an out-of-distribution image, using this as a measure of uncertainty.

Furthermore, a considerable body of work on OOD detection has taken measures to ensure that in-distribution and out-distribution data are separate. These works, including [4, 7, 10, 18, 15, 13], have employed entirely different datasets as in-distribution and out-distribution sources or used distinct classes as in-distribution and out-distribution data. It is our viewpoint that neither of these scenarios accurately reflects real-world operational conditions.

3.Deep Nearest Neighbor for OOD detection

In this section, we outline our approach to out-of-distribution (OOD) detection, utilizing deep k-Nearest Neighbor (KNN) for this purpose. This approach is visually represented in Figure 1 and, at a high level, falls under the category of distance-based methods.

Distance-based methods rely on feature embeddings obtained from a model and operate with the assumption that OOD test samples will exhibit considerable separation from the in-distribution (ID) data. In previous distance-based OOD detection methods, parametric density estimation was employed, and the feature embedding space was modeled as a mixture of multivariate Gaussian distributions (as demonstrated by Lee et al., 2018). However, this approach relies on a strong distributional assumption about the learned feature space, which may not always hold.

In this paper, we deviate from this approach and explore the effectiveness of non-parametric density estimation using nearest neighbors for OOD detection. Despite its simplicity, the KNN approach has not been systematically explored or compared in most contemporary OOD detection research. Specifically, we calculate the k-th nearest neighbor distance between the embedding of each test image and the training set. We employ a straightforward threshold-based criterion to determine if an input is OOD or not. Crucially, we use the normalized penultimate feature vector (denoted as $z = \phi(x)/\|\phi(x)\|_2$), where $\phi : X \rightarrow \mathbb{R}^m$ is a feature encoder, for OOD detection. During testing, we compute the normalized feature vector $z*$ for a test sample $x*$, and calculate the Euclidean distances to the embedding vectors zi in the training set, resulting in a reordering of these vectors based on the increasing distance. The decision function for OOD detection is defined $G(z*; k) = 1\{-r\_k(z*) \geq \lambda\}$, where $r\_k(z*) = \|z* - z(k)\|_2$ represents the distance to the k-th nearest neighbor (k-NN), and $1\{\cdot\}$ is the indicator function. Importantly, the threshold λ is typically chosen such that a high fraction of ID data (e.g., 95%) is correctly classified, and it does not rel on OOD data.

The presented approach is summarized in Algorithm 1. The KNN-based OOD detection method offers several notable advantages:

1. **Distributional Assumption-Free**: It does not impose any distributional assumptions on the underlying feature space, providing greater flexibility and applicability, even in cases where the feature space doesn't conform to Gaussian mixtures.

2. **OOD-Agnostic**: The testing procedure does not require information about the OOD data. The distance threshold is solely estimated using ID data.

3. **Ease of Use**: Modern implementations of approximate nearest neighbor search allow efficient processing, even with large databases containing billions of images, unlike the Mahalanobis distance, which requires potentially numerically unstable calculations.

4. **Model-Agnostic**: The testing procedure is applicable to various model architectures, including Convolutional Neural Networks (CNNs) and more recent Transformer-based Vision Transformer (ViT) models. Furthermore, the KNN approach is agnostic to the training procedure and is compatible with models trained using different loss functions, such as cross-entropy loss and contrastive loss.

The effectiveness of the KNN-based OOD detection approach is demonstrated in Section 4 of the paper, showing its potential for enhancing OOD detection in machine learning models.

4. Our method

Building upon our hypothesis that a lower out-of-distribution (OOD) score enhances reliability, we propose a practical approach to adapt a vision system rather than outright discarding OOD data. To put this into action, we establish a framework for adaptation within the context of an existing vision task, specifically object detection using a YOLOv4 network.

Many commercially available cameras provide access to several adjustable parameters that can influence the visual characteristics of captured images. Examples of such parameters include saturation, contrast, and exposure. These camera parameters naturally lend themselves to adaptation for various imaging conditions.

Our method revolves around the straightforward concept of adjusting camera parameters to minimize the OOD score. Intuitively, our approach can be deployed in two ways:

- First, when an image is identified as OOD, we adapt the camera parameters to reduce the OOD score.

- Alternatively, we can continuously adapt the camera parameters to maintain a minimized OOD score.

Although most cameras come with built-in features like autoexposure, these may prove inadequate for achieving the desired results. As illustrated in Figure 1, we provide two images of the same scene to emphasize this point. In one image, the camera's default settings fail to produce an image suitable for object detection. However, with carefully adjusted parameters, we obtain a more suitable image. It's worth noting that the yellow bounding box in the hand-tuned image indicates that YOLOv4 successfully detects the dog, a task that is not achievable with the camera's default settings.

5. Theoretical Justification

In this section, we delve into a theoretical analysis of employing the k-Nearest Neighbor (KNN) method for out-of-distribution (OOD) detection. By modeling the KNN within the feature space, our theoretical framework accomplishes two critical objectives: (1) it establishes a direct connection with our practical approach, which also operates in the feature space, and (2) it complements our experimental findings by considering the universality of OOD data. Our primary goal here is to analyze the average performance of our algorithm while remaining OOD-agnostic and training-agnostic.

Setup:

We view the OOD detection task as a specialized binary classification problem, where the negative samples (OOD) are only available during testing. The inputs come from the feature embedding space Z, and the labeling set is denoted as $G = \{0 \text{ (OOD)}, 1 \text{ (ID)}\}$. During inference, the testing set $\{(z_i, g_i)\}$ is drawn independently and identically distributed (i.i.d.) from P_{ZG}.

We introduce the marginal distribution on Z as P. To model the potential occurrence of both ID and OOD data during testing, we adopt the Huber contamination model, which represents the distribution as a combination of P_{in} and P_{out}. Here, P_{in} and P_{out} are the underlying distributions of feature embeddings for ID and OOD data, respectively, and ε is a constant that controls the fraction of OOD samples during testing. We use $p_{in}(z_i)$ and $p_{out}(z_i)$ as the probability density functions, where $p_{in}(z_i) = p(z_i \mid g_i = 1)$ and $p_{out}(z_i) = p(z_i \mid g_i = 0)$.

A significant challenge in OOD detection, both in practice and in theoretical analysis, is the lack of knowledge about the OOD distribution, which can vary universally outside the realm of ID data. To address this, we maintain a general analysis framework that acknowledges our limited prior information about OOD. Consequently, we model OOD data as having an equal chance of appearing outside the high-density region of ID data, described as $p_{out}(z) = c_0 * 1\{p_{in}(z) < c_1\}$.

Without access to an oracle that provides information about the underlying density function, our method relies on KNN as a distance measure that acts as a probability density estimator and, subsequently, establishes the decision boundary based on it. Specifically, the hypothesis class of KNN is defined as $\{h : h_{\lambda,k,Z_n}(z_i) = 1\{-r_k(z_i) \geq \lambda\}\}$, where $r_k(z_i)$ is the distance to the k-th nearest neighbor.

Main Result:

We demonstrate that our KNN-based OOD detector can reject inputs that are equivalent to the estimated Bayesian binary decision function. A small KNN distance $r_k(z_i)$ directly corresponds to a high probability of being classified as ID, and vice versa. This result is encapsulated in the following theorem.

Theorem 5.1: Under the specified setup, if $\hat{p}_{out}(z_i) = \hat{c}_0 * 1\{\hat{p}_{in}(z_i ; k, n) < \beta\varepsilon\hat{c}_0 (1-\beta)(1-\varepsilon)\}$, and $\lambda = -m-1 * q(1-\beta)(1-\varepsilon)k / (\beta\varepsilon cbn\hat{c}_0)$, then we have $1\{-r_k(z_i) \geq \lambda\} = 1\{\hat{p}(g_i = 1|z_i) \geq \beta\}$.

In experiments, as it is challenging to simulate universal OOD data, we approximate it by using a diverse yet finite collection of datasets. Our theoretical framework thus complements our practical findings and captures the universality of OOD data. It's important to note that β does not necessarily have to be 1/2 for the Bayesian classifier to be optimal; β can take any value greater than $(1-\varepsilon)c_1 / ((1-\varepsilon)c_1+\varepsilon c_0)$ when $\varepsilon c_0 \geq (1 - \varepsilon)c_1$.

6. Conclusion

This paper has demonstrated the feasibility of training a normalizing flow model for out-of-distribution (OOD) detection on a comprehensive and varied dataset like COCO. Furthermore, we conducted a real-world experiment that illustrates the practical applicability of a normalizing flow OOD detector as a quality metric for vision-related tasks. By optimizing camera parameters based on the output from our normalizing flow OOD model, we achieved an average improvement of 3 to 4 percentage points in the reliability of YOLOv4, specifically in terms of object detection performance.

7. Related Work

The issue of neural networks exhibiting overconfidence in out-of-distribution (OOD) data was initially brought to attention in the work of Nguyen et al. (2015). This concern has led to various research directions, including:

1. **Scoring Functions:** Researchers have developed scoring functions for OOD detection, such as the OpenMax score, maximum softmax probability, ODIN score, deep ensembles, Mahalanobis distance-based score, energy score, activation rectification (ReAct), gradient-based score, and ViM score. Notably, the need for evaluating OOD detection methods in real-world scenarios has been highlighted, as methods developed for smaller datasets like CIFAR may not effectively generalize to larger benchmarks like ImageNet. Prior to our work, the non-parametric nearest neighbor approach had not been explored for OOD detection, and our study fills this gap, demonstrating its effectiveness on multiple OOD detection benchmarks.

2. **Training-Time Regularization:** Some approaches address OOD detection through regularization techniques applied during model training. These methods encourage models to make predictions with a uniform distribution or higher energies for OOD data. Some require auxiliary OOD data, while others, like VOS, synthesize virtual outliers for regularization.

3. **Representation Learning:** Recent works have delved into the role of representation learning in OOD detection. For instance, CSI investigates data augmentations beneficial for OOD detection, while other works explore the effectiveness of applying multiview contrastive losses like SimCLR and SupCon for OOD detection. These methods often make distributional assumptions about the feature space.

Our approach differs fundamentally from previous methods in OOD detection, as we utilize KNN, a non-parametric method that doesn't rely on prior assumptions about the ID distribution. Our approach demonstrates superior performance compared to methods like SSD and is practical for real-world use.

KNN has been explored for anomaly detection, aiming to identify abnormal input samples within a single class. In contrast, our focus is OOD detection, which involves multi-class classification for ID data. While KNN-based anomaly detection for tabular data has been studied, its potential for OOD detection in deep neural networks remains underexplored. Our work contributes new empirical insights and theoretical analysis, demonstrating the efficacy of the KNN-based approach for OOD detection.

8. Reference

Certainly, here's a rephrased version of your provided text:

1. The Intel RealSense Depth Camera D435 is a device used for depth sensing and computer vision applications. More information can be found on the official Intel RealSense website: [Intel RealSense Depth Camera D435](https://www.intelrealsense.com/depth-camerad435/).

2. The YOLOv4 (You Only Look Once version 4) is an object detection model known for achieving an optimal balance between speed and accuracy. For detailed information, you can refer to the original research paper: [YOLOv4: Optimal Speed and Accuracy of Object Detection](https://arxiv.org/abs/2004.10934).

3. If you're interested in measure theory, you may want to explore the book "Measure Theory, Volume 1" by V.I. Bogachev and M.A.S. Ruas, available from Springer (2007).

4. Image anomaly detection with generative adversarial networks is an area of research. You can find more insights on this topic in the paper titled

"Image Anomaly Detection with Generative Adversarial Networks" by L. Deecke, R. Vandermeulen, L. Ruff, S. Mandt, and M. Kloft. It's included in the proceedings of the Machine Learning and Knowledge Discovery in Databases conference.

5. Real NVP (Real Normalizing Flow) is a technique for density estimation. If you want to delve deeper into this subject, you can refer to the paper "Density Estimation Using Real NVP" by L. Dinh, J. Sohl-Dickstein, and S. Bengio, available [here](https://arxiv.org/abs/1605.08803).

6. The uncertainty in deep neural networks is a subject of interest. To explore this further, you can read a comprehensive survey on the topic titled "A Survey of Uncertainty in Deep Neural Networks." It's available at [this link](https://arxiv.org/abs/2107.03342).

7. Generalized ODIN (Out-of-Distribution Detector in Neural networks) is a method for detecting out-of-distribution images without the need to learn from specific out-of-distribution data. More details on this technique can be found in the proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).

8. Adam is an optimization method designed for stochastic optimization tasks. For in-depth information about this method, you can refer to the original research paper titled "Adam: A Method for Stochastic Optimization."

9. Unfortunately, the provided text seems to be incomplete, as it ends abruptly. If you have further details or text to rephrase, please provide them, and I'd be happy to assist.