Deep SAD (Semi-supervised Anomaly Detection) using labeled and unlabeled data

1. Introduction

What is anomaly detection in general..

Anomaly detection is generally treated as an unsupervised learning problem. The approach consists of having a model learn the normal distribution from a set of given instances in the problem domain, and then, at detection time, use some form of a cross entropy loss function in order to detect anomalies.

There are a few issues with this approach. One of the key issues is that the mentioned approach does not utilize any known labeled anomalous data that may be available. The anomaly detection process is generally treated the same way as outlier detection. The issue with this is that *all detected anomalies are usually grouped together.*

At the same time, there do exist some semi-supervised approaches to anomaly detection which use labeled samples, but they are limited to merely including labeled normal samples. Very few *general* techniques exist which take advantage of labeled anomalies. In this article, we will discuss **Deep SAD [1]** which is an end-to-end generalized technique in deep semi-supervised learning (SSL) for anomaly detection.

Lastly, we will talk about comparative benchmarks with MNIST, Fashion-MNIST, and CIFAR-10 data sets.

II. High Level Comparison with Existing Approaches

There have been many techniques used for 1 2 3

Some important \_\_\_ techniques are but they dont \_\_\_

Research on deep semi-supervised learning has typically focused on classifiers which work under the cluster assumption.

Cluster assumption is \_\_\_

Using cluster assumption does not work because its assumptions are only based on the normal class. It importantly doesn’t explicitly consider the fact that distribution of the anomaly class may not be predictable ahead of time. So, finding anomalies by looking at any patterns outside of the normal class would not work.

Semi-supervised anomaly detection approaches must find a compact description of the normal class while correctly discriminating labeled anomalies

Deep SAD is an end-to-end deep method for general semi-supervised Anomaly Detection. The key things it introduces are:

* A generalization of unsupervised Deep SVDD method (Ruff, et al., 2018)
* Information theoretic framework for deep anomaly detection which can serve as an interpretation of our Deep SAD method and similar approaches
* Bench marking and proposal of evaluation methods for deep anomaly detection problems

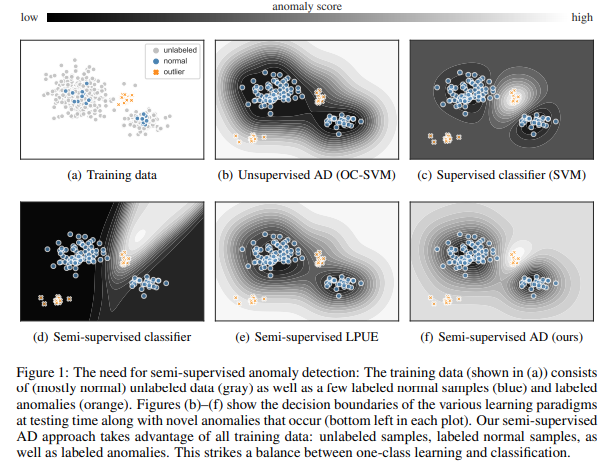
III. Information Theoretic Framework – contextualizing Deep SAD

In **supervised classification** where one has input variable X, latent variable Z, and output variable Y, the well known *information bottleneck principle* provides an explanation for representation learning as trade-off between finding a minimal compression Z of the input variable X while retaining informativeness of Z for predicting label Y. In short, supervised deep learning seeks to minimze the mutual information I(X; Z) between the input X and latent representation Z while maximizing mututal infromation I(Z; Y) between Z and classification task Y.

**Unsupervised learning** -- Infomax principle – objective is to maximize the mutual information I(X;Z) between the data X and its latent representation Z. This is a prevalent idea in autoencoders which are heavily used for anomaly detection.

Infomax principle is applied with autoencoders.

Based on all these techniques, Fig. 1 can be explained.



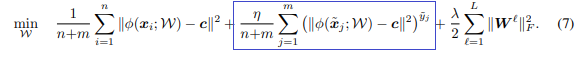
IV. Deep SAD

The objective of Deep SVDD is to train the neural network φ to learn a transformation that minimizes the volume of a data-enclosing hypersphere in output space Z centered on a predetermined point c

One-Class Deep SVDD objective is 

Penalizing mean squared distance of the mapped samples to the hypersphere center c forces the network to extract those common factors of variation that are the most stable within the dataset.

SECTION 3.2



For the labeled data, we introduce a new loss term that is weighted via the hyperparameter η > 0 which controls the balance between the labeled and the unlabeled term. Setting η > 1 puts more emphasis on the labeled data whereas η < 1 emphasizes the unlabeled data

Pg. 5 “We define the Deep SAD an”

V. Benchmarks and Comparison

VI. Conclusion

References:

[1] IEEE reference for the main paper

<https://arxiv.org/pdf/1906.02694.pdf>