Deep Nearest Neighbor Anomaly Detection

Liron Bergman [*1] **Niv Cohen** [*1] **Yedid Hoshen** [1]

Abstract

Nearest neighbors is a successful and long-standing technique for anomaly detection. Significant progress has been recently achieved by self-supervised deep methods (e.g. RotNet). Self-supervised features however typically under-perform Imagenet pre-trained features. In this work, we investigate whether the recent progress can indeed outperform nearest-neighbor methods operating on an Imagenet pretrained feature space. The simple nearest-neighbor based-approach is experimentally shown to outperform self-supervised methods in: accuracy, few shot generalization, training time and noise robustness while making fewer assumptions on image distributions.

1. Introduction

Agents interacting with the world are constantly exposed to a continuous stream of data. Agents can benefit from classifying particular data as anomalous i.e. particularly interesting or unexpected. Such discrimination is helpful in allocating resources to the observations that require it. This mechanism is used by humans to discover opportunities or alert of dangers. Anomaly detection by artificial intelligence has many important applications such as fraud detection, cyber intrusion detection and predictive maintenance of critical industrial equipment.

In machine learning, the task of anomaly detection consists of learning a classifier that can label a data point as normal or anomalous. In supervised classification, methods attempt to perform well on normal data whereas anomalous data is considered noise. The goal of an anomaly detection methods is to specifically detect extreme cases, which are highly variable and hard to predict. This makes the task of anomaly detection challenging (and often poorly specified).

The three main settings for anomaly detection are: supervised, semi-supervised and unsupervised. In the *supervised* setting, labelled training examples exist for normal and anomalous data. It is therefore not fundamentally different from other classification tasks. This setting is also too restrictive for many anomaly detection tasks as many anomalies of interest have never been seen before e.g. the emergence of new diseases. In the more interesting *semi-supervised* setting, all training images are normal with no included anomalies. The task of learning a normal-anomaly classifier is now one-class classification. The most difficult setting is *unsupervised* where an unlabelled training set of both normal and anomalous data exists. The typical assumption is that the proportion of anomalous data is significantly smaller than normal data. In this paper, we deal both with the semi-supervised and the unsupervised settings. Anomaly detection methods are typically based on distance, distribution or classification. The emergence of deep neural networks has brought significant improvements to each category. In the last two years, deep classification-based methods have significantly outperformed all other methods, mainly relying on the principle that classifiers that were trained to perform a certain task on normal data will perform this task well on unseen normal data, but will fail on anomalous data, due to poor generalization on a different data distribution.

In a recent paper, Gu et al. (2019) demonstrated that a K nearest-neighbours (kNN) approach on the raw data is competitive with the state-of-the-art methods on tabular data. Surprisingly, kNN is not used or compared against in most current image anomaly detection papers. In this paper, we show that although kNN on raw image data does not perform well, it outperforms the state of the art when combined with a strong off-the-shelf generic feature extractor. Specifically, we embed every (train and test) image using an Imagenet-pretrained ResNet feature extractor. We compute the K nearest neighbor (KNN) distance between the embedding of each test image and the training set, and use a simple threshold-based criterion to determine if a datum is anomalous.

We evaluate this baseline extensively, both on commonly used datasets as well as datasets that are quite different from Imagenet. We find that it has significant advantages over existing methods: i) higher than state-of-the-art accuracy ii) extremely low sample complexity iii) it can utilize

[*]Equal contribution [1]School of Computer Science and Engineering, The Hebrew University of Jerusalem, Israel. Correspondence to: Yedid Hoshen <yedid.hoshen@mail.huji.ac.il>.