

例えば、画像は回転不変であり、任意のサイズであることが可能である。

もう一つの一般的なアプローチは、1クラスSVM (Scholkopf et al., 2000)と関連するSVDD (Tax & Duin, 2004)である。SVDDは、少なくともある割合の正常データ点を含む最小体積球を当てはめるものと見なすことができる。この方法は特徴空間に対して非常に敏感であるため、効果的な特徴空間を学習するためにカーネル法を用いた。

異常検出にkNNを使用することは新しい手法ではないが、最近の画像異常検出の研究ではあまり使用されず、比較もされていない。我々の目的は、この単純だが非常に効果的で一般的な画像異常検知手法を認知させることである。そのシンプルさ、ロバスト性、低いサンプル複雑性、一般性から、全ての新しい研究はこのシンプルな手法と比較されるべきだと我々は考えている。

2. Previous Work

プリディープ学習法 異常検出のための2つの古典的なパラダイムは、再構成ベースと分布ベースである。再構成ベースの手法は、学習セットを使用して、正常データを効果的な方法で表現する基底関数のセットを学習する。テスト時には、学習した基底関数を用いて新しいサンプルの再構成を試みる。この方法では、正常なデータはうまく再構成されるが、異常なデータは再構成されないと仮定する。再構成コストを閾値化することで、サンプルが正常か異常かに分類される。さまざまな基底関数の選択には、他のサンプルの疎な組み合わせ (例: kNN) (Eskin et al., 2002)、主成分 (Jolliffe, 2011; Candès et al., 2011)、K-means (Hartigan & Wong, 1979) などがある。再構成メトリックには、ユークリッド距離、 L_1 距離、あるいはSSIM (Wang et al.) 再構成に基づく手法の主な弱点は、i) 識別基底関数の学習が困難である ii) 効果的な類似性尺度を見つけることが自明でないこと、である。半教師付き分布ベースのアプローチは、正規データの確率密度関数 (PDF) を学習しようとするものである。新しいサンプルが与えられると、その確率が評価され、確率がある閾値より低ければ異常と判定される。このような手法には次のようなものがある: ガウス混合 (GMM) などのパラメトリック・モデル。非パラメトリック手法には、Kernel Density Estimation (Latecki et al., 2007)やkNN (Eskin et al., 2002)がある (我々は再構成ベースも考慮する)。

分布手法の主な弱点は、高次元データの密度推定が難しいことである。

ディープネットワークによる古典的手法の補強: ディープニューラルネットワークの成功は、ディープに学習された特徴を古典的な手法と組み合わせる研究を促した。PCA法はディープオートエンコーダーに拡張され (Yang et al., 2017)、再構成コストはディープ知覚損失に拡張された (Zhang et al.) GANは画像における再構成のための基底関数としても用いられた。分布モデルを改善する1つのアプローチ (Zong et al., 2018) は、まずデータを意味的な低次元空間に埋め込むことを学習し、次にGMMなどの標準的な手法を用いてその分布をモデル化することである。SVDDはRuffら (2018)によって、カーネル法の優れた代替法として深層特徴を学習するために拡張された。この方法は「モード崩壊」の問題に悩まされており、後続の研究の対象となっている。この論文で調査されたアプローチは、古典的なkNNが深く学習された特徴で拡張されているため、このカテゴリーに属すると見なすことができる。

自己教師付きディープメソッド: 深層表現を学習するために監視を使用する代わりに、自己教師ありの方法は、データを入手するのが無料であるか、少なくとも非常に安価である補助的なタスクを解決するためにニューラルネットワークを訓練する。自己教師ありの表現は、通常、Imagenetのような大規模な教師ありデータセットから学習されたものを下回ることに注意すべきである。高品質な画像特徴を学習するための補助的なタスクとしては、ビデオフレーム予測 (Mathieu et al., 2016)、画像カラー化 (Zhang et al., 2016; Larsson et al., 2016)、パズル解法 (Nooroozi & Favaro, 2016) などがある。最近、Gidarisら (2018)は一連の画像処理変換 (画像軸を中心に0、90、180、270度回転)を用い、真の画像の向きを予測した。彼らはこれを用いて高品質の画像特徴を学習した。Golan & El-Yaniv (2018)は、画像の異常を検出するために同様の画像処理タスク予測を用いた。この手法は、異常クラスからの画像検出において良好な性能を示した。この手法の性能はHendrycksら (2019)によって改善され、一方Bergman & Hoshen (2020)によってopenset分類と組み合わせられ、表形式データに拡張された。本研究では、画像異常検出タスクにおいて、自己教師付き手法が、強力な汎用特徴抽出器を用いた、より単純なkNNベースの手法を下回ることを示す。

本稿のもう一つの貢献は、深層学習コミュニティからほとんど注目されていなかったタスクである、画像グループの異常検出へのkNNの新しい適応を提示することである。

我々は、画像異常検出のための単純なK最近傍 (kNN) ベースの手法を研究する。この手法をDeep Nearest Neighbors (DN2)と呼ぶ。

very strong external feature extractors, at minimal cost iv) it makes few assumptions on the images e.g. images can be rotation invariant, and of arbitrary size v) it is robust to anomalies in the training set i.e. it can handle the unsupervised case (when coupled with our two-stage approach) vi) it is plug and play, does not have a training stage.

Another contribution of our paper is presenting a novel adaptation of kNN to image group anomaly detection, a task that received scant attention from the deep learning community.

Although using kNN for anomaly detection is not a new method, it is not often used or compared against by most recent image anomaly detection works. Our aim is to bring awareness to this simple but highly effective and general image anomaly detection method. We believe that every new work should compare to this simple method due to its simplicity, robustness, low sample complexity and generality.

2. Previous Work

Pre-deep learning methods: The two classical paradigms for anomaly detection are: reconstruction-based and distribution-based. Reconstruction-based methods use the training set to learn a set of basis functions, which represent the normal data in an effective way. At test time, they attempt to reconstruct a new sample using the learned basis functions. The method assumes that normal data will be reconstructed well, while anomalous data will not. By thresholding the reconstruction cost, the sample is classified as normal or anomalous. Choices of different basis functions include: sparse combinations of other samples (e.g. kNN) (Eskin et al., 2002), principal components (Jolliffe, 2011; Candès et al., 2011), K-means (Hartigan & Wong, 1979). Reconstruction metric include Euclidean, L_1 distance or perceptual losses such as SSIM (Wang et al., 2004). The main weaknesses of reconstruction-based methods are i) difficulty of learning discriminative basis functions ii) finding effective similarity measures is non-trivial. Semi-supervised distribution-based approaches, attempt to learn the probability density function (PDF) of the normal data. Given a new sample, its probability is evaluated and is designated as anomalous if the probability is lower than a certain threshold. Such methods include: parametric models e.g. mixture of Gaussians (GMM). Non-parametric methods include Kernel Density Estimation (Latecki et al., 2007) and kNN (Eskin et al., 2002) (which we also consider reconstruction-based). The main weakness of distributional methods is the difficulty of density estimation for high-dimensional data. Another popular approach is one-class SVM (Scholkopf et al., 2000) and related SVDD (Tax & Duin, 2004). SVDD can be seen as fitting the minimal volume sphere that includes at least a certain percentage of the normal data points. As this method

is very sensitive to the feature space, kernel methods were used to learn an effective feature space.

Augmenting classical methods with deep networks: The success of deep neural networks has prompted research combining deep learned features to classical methods. PCA methods were extended to deep auto-encoders (Yang et al., 2017), while their reconstruction costs were extended to deep perceptual losses (Zhang et al., 2018). GANs were also used as a basis function for reconstruction in images. One approach (Zong et al., 2018) to improve distributional models is to first learn to embed data in a semantic, low dimensional space and then model its distribution using standard methods e.g. GMM. SVDD was extended by Ruff et al. (2018) to learn deep features as a superior alternative for kernel methods. This method suffers from a "mode collapse" issue, which has been the subject of followup work. The approach investigated in this paper can be seen as belonging to this category, as classical kNN is extended with deep learned features.

Self-supervised Deep Methods: Instead of using supervision for learning deep representations, self-supervised methods train neural networks to solve an auxiliary task for which obtaining data is free or at least very inexpensive. It should be noted that self-supervised representation typically underperform those learned from large supervised datasets such as Imagenet. Auxiliary tasks for learning high-quality image features include: video frame prediction (Mathieu et al., 2016), image colorization (Zhang et al., 2016; Larsson et al., 2016), and puzzle solving (Noroozi & Favaro, 2016). Recently, Gidaris et al. (2018) used a set of image processing transformations (rotation by 0, 90, 180, 270 degrees around the image axis), and predicted the true image orientation. They used it to learn high-quality image features. Golan & El-Yaniv (2018), have used similar image-processing task prediction for detecting anomalies in images. This method has shown good performance on detecting images from anomalous classes. The performance of this method was improved by Hendrycks et al. (2019), while it was combined with openset classification and extended to tabular data by Bergman & Hoshen (2020). In this work, we show that self-supervised methods underperform simpler kNN-based methods that use strong generic feature extractors on image anomaly detection tasks.

3. Deep Nearest-Neighbors for Image Anomaly Detection

We investigate a simple K nearest-neighbors (kNN) based method for image anomaly detection. We denote this method, Deep Nearest-Neighbors (DN2).