3.1. 半教師付き異常検知

DN2 は入力画像の集合 $X_{train} = x_1 , x_2 .. x_N$ を取ります。 半教師付き設定では、全ての入力画像が正常であると仮定します。DN2は、訓練済みの特徴抽出器Fを用いて、訓練セット全体から特徴を抽出する：

$$f_i = F(x_i) \tag{1}$$

本稿では、Imagenetデータセットで事前学習されたResNet特徴抽出器 を使用する。一見すると、この監視は強い要求であるように見えるかもしれないが、このような特徴抽出器は広く利用可能である。正常な画像や異常な画像がImagenetと特に密接に関連している必要はないことは、後で実験的に示す。

$f_{train} = f_1 , f_2 .. f_N$. 初期段階の後、埋め込みは保存され、学習セットの推論を償却することができる。

新しいサンプルyが異常かどうかを推論するために、我々はまずその特徴埋め込みを抽出する： $f_y = F(y)$. $f_y = F(y)$ # 次にそのkNN距離を計算し、異常スコアとして利用する：

$$d(y) = \frac{1}{k} \sum_{f \in N_k(f_y)} \|f - f_y\|^2 \tag{2}$$

$N_k(f_y)$ は、学習集合 F_{train} における f_y に最も近い k 個の埋め込みを表す。我々はユークリッド距離を使用することにした。これはディープネットワークによって抽出された特徴量に対してしばしば強力な結果をもたらすが、他の距離尺度も同様の方法で使用することができる。距離d(y)が閾値より大きいかどうかを検証することで、画像yが正常か異常かを判定する。

3.2. 教師なし異常検知

完全に教師なしな場合、全ての入力画像が正常であると仮定することはできなくなり、代わりに、ごく一部の入力画像のみが異常であると仮定する。このより困難な設定に対処するために（そして教師なし異常検出に関する先行研究にも沿う）、我々はまず入力画像に対してクリーニング段階を行うことを提案する。特徴抽出段階の後、各入力画像と残りの入力画像との間のkNN距離を計算する。異常画像は低密度の領域にあると仮定して、kNN距離が最大の画像の一部を除去する。この割合は、推定される異常な入力画像の割合よりも大きくなるように選ぶべきである。DN2が非常に少ない学習画像しか必要としないことは、後の実験で示される。従って、除去する画像の割合に非常に積極的になり、正常である可能性が最も高い画像だけを残すことができる（実際にはトレーニング画像の50%を除去する）。異常が疑われる入力画像を除去した後、画像は正常な画像の割合が非常に高いと仮定する。したがって、半教師ありの場合と全く同じように進めることができる。

3.3. グループ画像の異常検出

グループ異常検知は、入力サンプルが画像の集合からなる設定に取り組む。特定の組み合わせは重要だが、順番は重要ではない。集合内の各画像は個々には正常であるが、集合全体としては異常である可能性がある。例として、M個の画像からなる正常な集合を想定しよう。点（画像単位）の異常検出器を学習させた場合、点的に異常な画像を含む異常集合を検出することができる。しかし、あるクラスからの複数の画像を含み、別のクラスからの画像を含まない異常集合は、全ての画像が個々に正常であるため、正常集合として分類される。以前、画像におけるグループ異常検出に取り組むために、いくつかのディープオートエンコーダ手法が提案された（例えばDOroら（2019））。このような手法は、i）高いサンプル複雑性 ii）再構成メトリックに対する感度 iii）グループに対する感度の潜在的欠如、という複数の欠点に苦しんでいる。我々は効果的なkNNベースのアプローチを提案する。提案手法は、集合内の画像の全ての特徴に対して無秩序にプール（我々は平均化を選んだ）することで集合を埋め込む：

1. グループg内の全画像から特徴抽出 $f^i_g = F(x^i_g)$ # 4.1.
2. グループ全体にわたる特徴の無秩序なプール：
   $f_g = \frac{\sum_i f_g^i}{number\ of\ images}$

上述のグループ特徴を抽出した後、DN2を用いて異常の検出に進む。

4. Experiments

このセクションでは、上述の単純なkNNアプローチが、最先端の性能よりも優れた性能を達成することを示す、広範な実験を紹介する。この結論は、タスクやデータセットを超えて一般化される。我々はこの手法をノイズに対してより頑健に拡張し、教師なし設定に適用できるようにする。さらに、この方法をグループ異常検出に有効なように拡張する。

4.1. 単峰性異常検出

異常検知手法を評価するための最も一般的な設定は単峰性である。この設定では，1つのクラスを正常とし，他のクラスを異常とすることで，分類データセットが適応される . 正常な訓練セットは手法の訓練に使用され、全てのテストデータは手法の推論性能を評価するために使用される。先行研究と同様に、ROC曲線下面積（ROCAUC）を報告する。

3.1. Semi-supervised Anomaly Detection

DN2 takes a set of input images $X_{train} = x_1, x_2..x_N$. In the semi-supervised setting we assume that all input images are normal. DN2 uses a pre-trained feature extractor F to extract features from the entire training set:

$$f_i = F(x_i) \tag{1}$$

In this paper, we use a ResNet feature extractor that was pretrained on the Imagenet dataset. At first sight it might appear that this supervision is a strong requirement, however such feature extractors are widely available. We will later show experimentally that the normal or anomalous images do not need to be particularly closely related to Imagenet.

The training set is now summarized as a set of embeddings $F_{train} = f_1, f_2..f_N$. After the initial stage, the embeddings can be stored, amortizing the inference of the training set.

To infer if a new sample y is anomalous, we first extract its feature embedding: $f_y = F(y)$. We then compute its kNN distance and use it as the anomaly score:

$$d(y) = \frac{1}{k} \sum_{f \in N_k(f_y)} \|f - f_y\|^2 \tag{2}$$

$N_k(f_y)$ denotes the k nearest embeddings to f_y in the training set F_{train}. We elected to use the euclidean distance, which often achieves strong results on features extracted by deep networks, but other distance measures can be used in a similar way. By verifying if the distance $d(y)$ is larger than a threshold, we determine if an image y is normal or anomalous.

3.2. Unsupervised Anomaly Detection

In the fully-unsupervised case, we can no longer assume that all input images are normal, instead, we assume that only a small proportion of input images are anomalous. To deal with this more difficult setting (and inline with previous works on unsupervised anomaly detection), we propose to first conduct a cleaning stage on the input images. After the feature extraction stage, we compute the kNN distance between each input image and the rest of the input images. Assuming that anomalous images lie in low density regions, we remove a fraction of the images with the largest kNN distances. This fraction should be chosen such that it is larger than the estimated proportion of anomalous input images. It will be later shown in our experiments that DN2 requires very few training images. We can therefore be very aggressive in the percentage of removed image, and keep only the images most likely to be normal (in practice we remove 50% of training images). After removal of the suspected anomalous input images, the images are now assumed to have a very high-proportion of normal images. We can therefore proceed exactly as in the semi-supervised case.

3.3. Group Image Anomaly Detection

Group anomaly detection tackles the setting where the input sample consists of a set of images. The particular combination is important, but not the order. It is possible that each image in the set will individually be normal but the set as a whole will be anomalous. As an example, let us assume normal sets consisting of M images, a randomly sampled image from each class. If we trained a point (per-image) anomaly detector, it will be able to detect anomalous sets containing pointwise anomalous images e.g. images taken from classes not seen in training. An anomalous set containing multiple images from one seen class, and no images from another will however be classified as normal as all images are individually normal. Previously, several deep autoencoder methods were proposed (e.g. DOro et al. (2019)) to tackle group anomaly detection in images. Such methods suffer from multiple drawbacks: i) high sample complexity ii) sensitivity to reconstruction metric iii) potential lack of sensitivity to the groups. We propose an effective kNN based approach. The proposed method embeds the set by orderless-pooling (we chose averaging) over all the features of the images in the set:

1. Feature extraction from all images in the group g, $f_g^i = F(x_g^i)$
2. Orderless pooling of features across the group: $f_g = \frac{\sum_i f_g^i}{number\ of\ images}$

Having extracted the group feature described above we proceed to detect anomalies using DN2.

4. Experiments

In this section, we present extensive experiments showing that the simple kNN approach described above achieves better than state-of-the-art performance. The conclusions generalize across tasks and datasets. We extend this method to be more robust to noise, making it applicable to the unsupervised setting. We further extend this method to be effective for group anomaly detection.

4.1. Unimodal Anomaly Detection

The most common setting for evaluating anomaly detection methods is unimodal. In this setting, a classification dataset is adapted by designating one class as normal, while the other classes as anomalies. The normal training set is used to train the method, all the test data are used to evaluate the inference performance of the method. In line with previous works, we report the ROC area under the curve (ROCAUC).