

図6. CIFAR10でSVDD（左）、Geometric（中央）、Imagenet事前学習（右）により学習された特徴量のt-SNEプロット（通常クラスはAirplane（上）、Automobile（下））。Imagenetで事前学習された特徴量は、正常クラス（黄色）と異常クラス（青色）を明確に分離していることがわかります。Geometricは、Airplaneについては不十分な特徴を学習し、Automobileについては妥当な特徴を学習する。Deep-SVDDはきれいな分離を可能にする特徴を学習しない。

$K \setminus N$ の比率で推論が高速化される。これは $\frac{N}{K}$ の比率で推論を高速化する。従って、学習特徴量をK個のクラスタにクラスタリングし、元の特徴量ではなくクラスタに対してkNNを実行することで、DN2を高速化することを提案する。表6にDN2の性能比較を示す。表6は、DN2とそのK-means近似の性能を、異なる平均数（2つの最近傍への距離の和を用いる）で比較したものである。わずかな精度の低下で、検索速度が大幅に低下することがわかる。

5.2. 事前学習された特徴量と自己教師付き特徴量の比較

事前訓練された特徴抽出器によるパフォーマンスの向上を理解するために、Deep-SVDD、Geometric、DN2（Imagenet上で事前訓練されたResnet50）によって抽出された正常なテスト特徴と異常なテスト特徴のt-SNEプロットを示します。上のプロットは、中程度の検出精度を達成した正常クラスのプロットであり、下のプロットは、高い検出精度を達成した正常クラスのプロットである。Deep-SVDDの正常クラスが異常クラスの中に散在していることがすぐに観察でき、その性能の低さを説明している。Geometricでは、正常クラスの特徴はもう少し局在していますが、正常領域の密度はまだ中程度にしか集中していません。

我々は、Geometricのかなり優れた性能は、それが実行する大規模なアンサンブル（72のオーグメントの組み合わせ）によって達成されると考えている。Imagenetが事前に学習した特徴量は、非常に強い局所性を保持していることがわかる。これはDN2の強力なパフォーマンスを説明する。

6. Discussion

異常検出のための一般的なパラダイム： 最近の論文（Golán & El-Yaniv(2018)など）は、自己監視のパラダイムを提唱している。本論文の結果は、別のパラダイムに強い証拠を与える： i) 曖昧に関連するデータセット上で利用可能な全ての監視を使用して一般的な特徴を学習する ii) 学習された特徴は、標準的な異常検出手法（例えばkNNやk-means）を使用できるように十分に一般的であることが期待される。事前学習されたパラダイムは、自己教師ありの手法よりもはるかに高速に導入でき、その他にも多くの利点がある。Imagenetとの類似性が全くない画像データに対しては、事前に訓練された特徴量を使用することはあまり効果的でない可能性がある。しかし、我々の実験では、Imagenetで事前学習された特徴量は、顕微鏡画像だけでなく、航空写真にも有効であることがわかった。

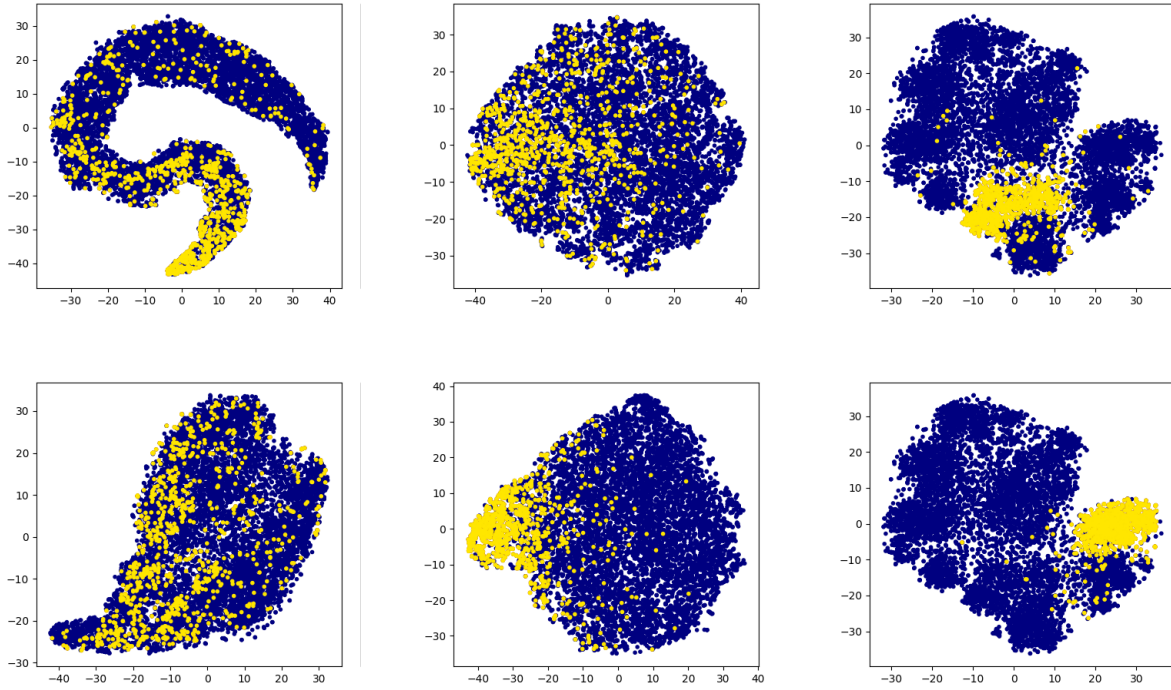


Figure 6. t-SNE plots of the features learned by SVDD (left), Geometric (center) and Imagenet pre-trained (right) on CIFAR10, where the normal class is Airplane (top), Automobile (bottom). We can see that Imagenet-pretrained features clearly separate the normal class (yellow) and anomalies (blue). Geometric learns poor features of Airplane and reasonable features on Automobile. Deep-SVDD does not learn features that allow clean separation.

time is using K-means clustering of the training features. This speeds up inference by a ratio of $\frac{N}{K}$. We therefore suggest speeding up DN2 by clustering the training features into K clusters and the performing kNN on the clusters rather than the original features. Tab. 6 presents a comparison of performance of DN2 and its K-means approximations with different numbers of means (we use the sum of the distances to the 2 nearest neighbors). We can see that for a small loss in accuracy, the retrieval speed can be reduced significantly.

5.2. Pretrained vs. self-supervised features

To understand the improvement in performance by pre-trained feature extractors, we provide t-SNE plots of normal and anomalous test features extracted by Deep-SVDD, Geometric and DN2 (Resnet50 pretrained on Imagenet). The top plots are of a normal class that achieves moderate detection accuracy, while the bottom plots are of a normal class that achieves high accuracy. We can immediately observe that the normal class in Deep-SVDD is scattered among the anomalous classes, explaining its lower performance. In Geometric the features of the normal class are a little more localized, however the density of the normal region is still only moderately concentrated. We believe that the

fairly good performance of Geometric is achieved by the massive ensembling that it performs (combination of 72 augmentations). We can see that Imagenet pretrained features preserve very strong locality. This explains the strong performance of DN2.

6. Discussion

A general paradigm for anomaly detection: Recent papers (e.g. Golan & El-Yaniv (2018)) advocated the paradigm of self-supervision, possibly with augmentation by an external dataset e.g. outlier exposure. The results in this paper, give strong evidence to an alternative paradigm: i) learn general features using all the available supervision on vaguely related datasets ii) the learned features are expected to be general enough to be able to use standard anomaly detection methods (e.g. kNN, k-means). The pretrained paradigm is much faster to deploy than self-supervised methods and has many other advantages investigated extensively in Sec. 4. We expect that for image data that has no similarity whatsoever to Imagenet, using pre-trained features may be less effective. That withstanding, in our experiments, we found that Imagenet-pretrained features were effective on aerial images as well as microscope images, while both settings