

ADTR: 特徴再構築付き異常検出トランスフォーマー

Zhiyuan You¹, Kai Yang², Wenhan Luo³, Lei Cui⁴, Yu Zheng¹, and Xinyi Le¹✉¹Shanghai Jiao Tong University, ²SenseTime Research
³Sun Yat-sen University, ⁴Tsinghua University

zhiyuanyou@foxmail.com, lexinyi@sjtu.edu.cn

要約. 異常サンプルの不足により、正常サンプルからの事前知識のみを用いた異常検出が注目されています。既存のCNNベースの 픽셀再構築アプローチには、2つの課題があります。まず、再構築のソースとターゲットは、区別不能なセマンティック情報を含む raw 픽셀値です。第二に、CNNは正常サンプルと異常を同様に良好に再構築するため、それらを区別するのが依然として困難です。本論文では、事前学習済み特徴の再構築にトランスフォーマーを適用する異常検出トランスフォーマー (ADTR) を提案します。事前学習済み特徴には区別可能なセマンティック情報が含まれています。また、トランスフォーマーの採用により、異常の再構築を適切に制限することで、再構築に失敗した場合に異常を容易に検出できるようになります。さらに、当手法を正常サンプルのみの場合と異常が利用可能な場合 (画像レベルと 픽셀レベルのラベル付き異常を含む) の両方に適応させるため、新たな損失関数を提案します。単純な合成異常や外部からの関連のない異常を追加することで、性能をさらに向上させることができます。MVTec-ADとCIFAR-10を含む異常検出データセットで広範な実験を実施した結果、当手法はすべてのベースラインよりも優れた性能を達成しました。

キーワード: 異常検出・トランスフォーマー・アテンションメカニズム。

1 Introduction

非監督型異常検出[4, 8, 15]は、正常なサンプルからの事前知識のみを使用して異常を特定することを目的としています。生産ラインにおける異常の極端な不足のため、異常検出はますます注目されています。

統計学的観点から、異常は正常サンプルの分布外れ値と見なすことができます。この設定では、AutoEncoder (AE)、Variational Auto-Encoder (VAE)、Generative Adversarial Network (GAN) などのCNNベースの再構築モデルが、正常サンプルの分布をモデル化するために一般的に採用されています[8, 13, 16, 19]。これらの手法は、一般化ギャップの仮定に基づき、正常なサンプルのみを使用してモデルを訓練します。つまり、再構築は正常なサンプルでは成功しますが、異常値では失敗します。異常検出は、サンプルとその再構築間の距離メトリクスを使用して行われます。

本研究は、上海市科学技术發展基金 (21SQBS01502) および中国国家自然科学基金 (62176152) の支援を受けています。

ADTR: Anomaly Detection Transformer with Feature Reconstruction

Zhiyuan You¹, Kai Yang², Wenhan Luo³, Lei Cui⁴, Yu Zheng¹, and Xinyi Le¹✉

¹Shanghai Jiao Tong University, ²SenseTime Research

³Sun Yat-sen University, ⁴Tsinghua University

zhiyuanyou@foxmail.com, lexinyi@sjtu.edu.cn

Abstract. Anomaly detection with only prior knowledge from normal samples attracts more attention because of the lack of anomaly samples. Existing CNN-based pixel reconstruction approaches suffer from two concerns. First, the reconstruction source and target are raw pixel values that contain indistinguishable semantic information. Second, CNN tends to reconstruct both normal samples and anomalies well, making them still hard to distinguish. In this paper, we propose Anomaly Detection TRansformer (ADTR) to apply a transformer to reconstruct pre-trained features. The pre-trained features contain distinguishable semantic information. Also, the adoption of transformer limits to reconstruct anomalies well such that anomalies could be detected easily once the reconstruction fails. Moreover, we propose novel loss functions to make our approach compatible with the normal-sample-only case and the anomaly-available case with both image-level and pixel-level labeled anomalies. The performance could be further improved by adding simple synthetic or external irrelevant anomalies. Extensive experiments are conducted on anomaly detection datasets including MVTec-AD and CIFAR-10. Our method achieves superior performance compared with all baselines.

Keywords: Anomaly Detection · Transformer · Attention Mechanism.

1 Introduction

Unsupervised anomaly detection [4,8,15] aims to identify anomalies using prior knowledge from only normal samples. Due to the extreme lack of anomalies in production lines, anomaly detection is attracting more and more interests.

From the view of statistics, anomalies may be seen as distribution outliers of normal samples. In this setting, CNN-based reconstruction models like Auto-Encoder (AE), Variational Auto-Encoder (VAE), and Generative Adversarial Network (GAN) are usually adopted to model the distribution of normal samples [8,13,16,19]. These methods train a model with only normal samples based on the assumption of generalization gap, which means that the reconstruction succeeds with only normal samples but fails with anomalies. The anomaly detection is performed with a distance metric between a sample and its reconstruction.

This work is sponsored by the Shanghai Foundation for Development of Science and Technology (21SQBS01502) and National Natural Science Foundation of China (62176152).

