ADTR: Anomaly Detection Transformer with Feature Reconstruction

Zhiyuan You[1], Kai Yang[2], Wenhan Luo[3], Lei Cui[4], Yu Zheng[1], and Xinyi Le[1]✉

[1]Shanghai Jiao Tong University, [2]SenseTime Research
[3]Sun Yat-sen University, [4]Tsinghua 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).