Unsupervised anomaly detection (AD) for image datasets
Case study: Thermal image of the fuel tank opening cap

Ph.D. student: LE Tuan

Thesis director(s): VU Hai Canh, BOULANGER Cherfi Zohra, NASSIM Boudaoud, PONCHET-DURUPT Amélie

Roberval laboratory
University of Technology of Compiegne
Alliance Sorbonne University

Email: tuan.le@utc.fr

27/03/2023

ALLIANCE SORBONNE UNIVERSITÉ

utc Recherche
Roberval

Table of content

Introduction

Introduction

- ✓ **Anomaly detection (AD)** involves identifying patterns in data that deviate from the expected behavior.

- ✓ The **application of AD** in various domains [1]: fraud detection in credit cards, health care insurance, intrusion detection in cybersecurity

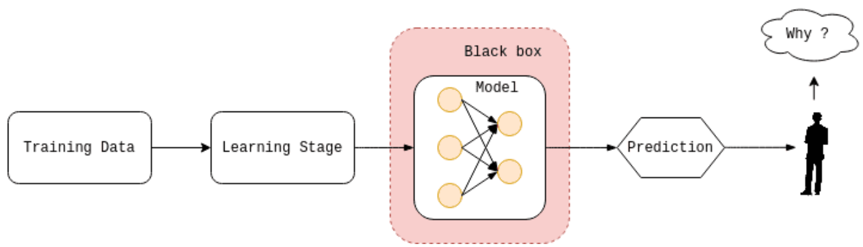
⇒ In this study, we focus on exploring **anomaly detection techniques for an image dataset in industrial systems**.

- ✓ After detecting anomalies, it is necessary to **provide the location of the anomalies**

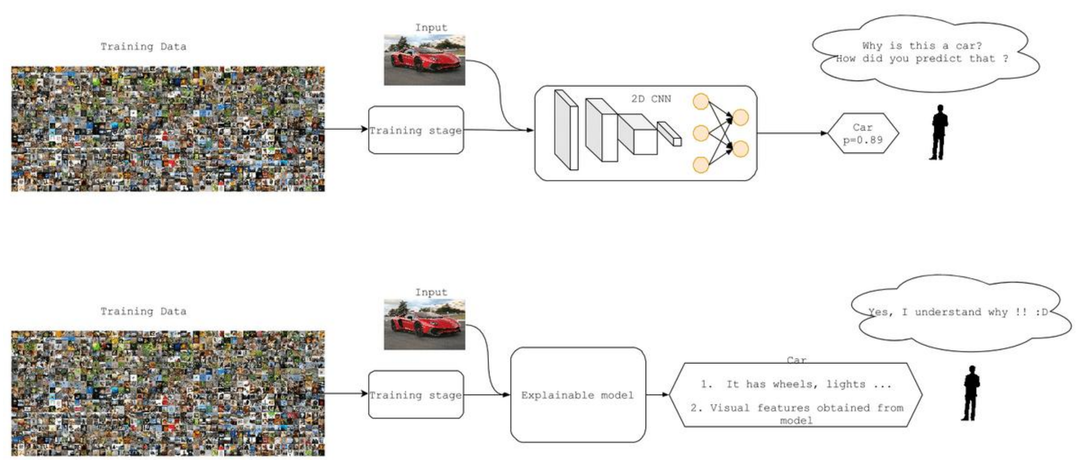
⇒ In this study, we also investigated several methods for this task (i.e. Explainable AI, U-transformer)

Explainable AI

In the context of anomaly, **explainable AI (XAI) is used to explain why a particular data point is considered an anomaly.**
By analyzing the data point and identifying the features or attributes that are most important in the model's decision to classify it as an anomaly.



Most machine learning models perform as black boxes.



Comparison of a deep learning and an explainable model.

Explainable AI: LIME

LIME, which stands for **Local Interpretable Model-Agnostic Explanations**, is a method for explaining the predictions made by machine learning models. It is particularly useful for explaining the decisions made by "black box" models, which are models that are difficult to interpret or understand.

(a) Original Image (b) Explaining *Electric guitar* (c) Explaining *Acoustic guitar* (d) Explaining *Labrador*

Figure 4: Explaining an image classification prediction made by Google's Inception neural network. The top 3 classes predicted are "Electric Guitar" ($p = 0.32$), "Acoustic guitar" ($p = 0.24$) and "Labrador" ($p = 0.21$)

(a) Husky classified as wolf (b) Explanation

Figure 11: Raw data and explanation of a bad model's prediction in the "Husky vs Wolf" task.

Source: [1] Ribeiro, Marco Tulio, Sameer Singh, and Carlos Guestrin. "" Why should I trust you?" Explaining the predictions of any classifier." Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 2016.

Explainable AI: Grad-CAM

Grad-CAM, or **Gradient Class Activation Mapping**, is a technique used to visualize **which parts of an image a deep learning model is looking at when making a prediction**.

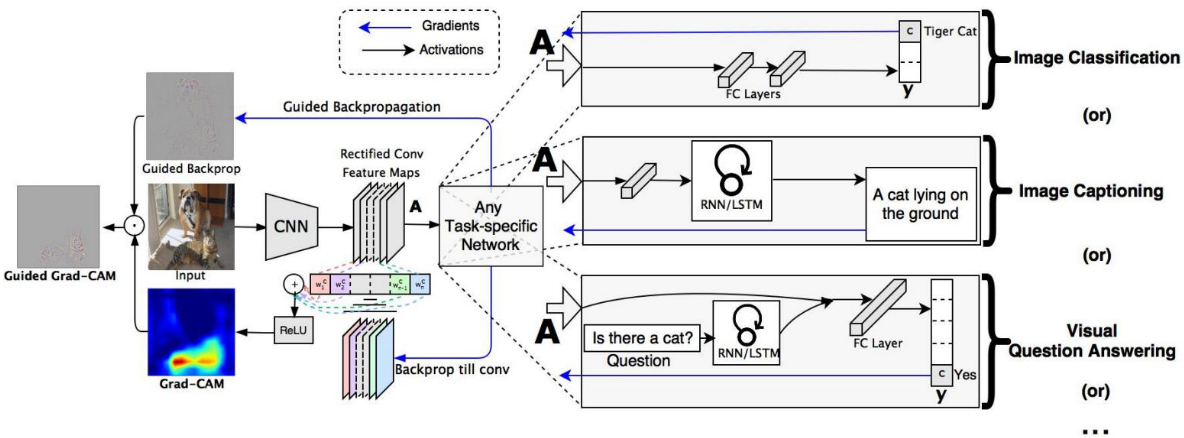
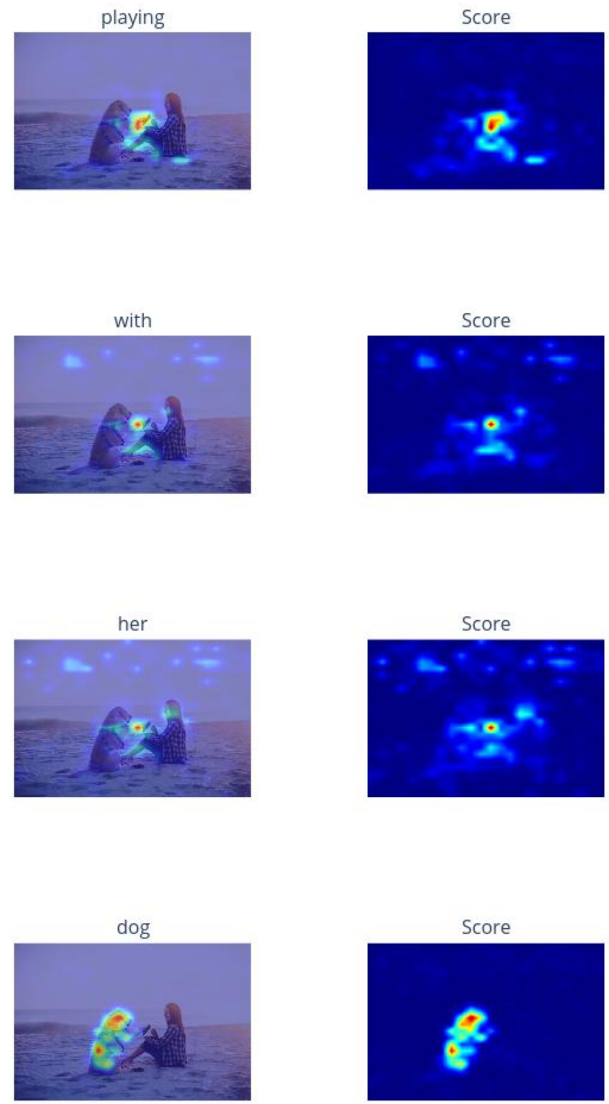


Fig. 2: Grad-CAM overview: Given an image and a class of interest (*e.g.*, 'tiger cat' or any other type of differentiable output) as input, we forward propagate the image through the CNN part of the model and then through task-specific computations to obtain a raw score for the category. The gradients are set to zero for all classes except the desired class (tiger cat), which is set to 1. This signal is then backpropagated to the rectified convolutional feature maps of interest, which we combine to compute the coarse Grad-CAM localization (blue heatmap) which represents where the model has to look to make the particular decision. Finally, we pointwise multiply the heatmap with guided backpropagation to get Guided Grad-CAM visualizations which are both high-resolution and concept-specific.



Source: [2] Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization

Model based on U-transformer

UTRAD is a new framework for anomaly detection in industrial defect detection.

- Pre-trained CNN: feature extraction.
- Skip connection: detect both structural and non-structural anomalies + reduce cost and memory.
- U-transformer: reconstruct input to output as an autoencoder.

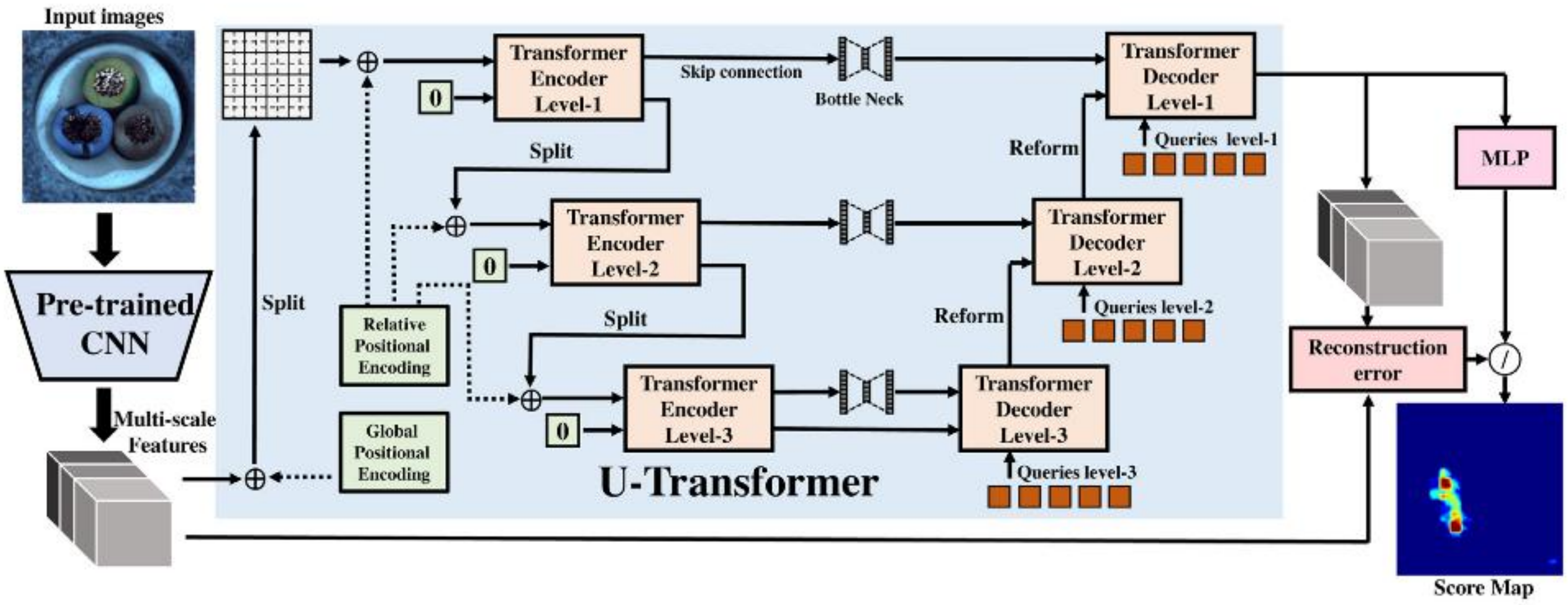


Fig. 1. The overall framework of the proposed anomaly detection and localization method. Firstly, The image is input to the pre-trained network to extract multi-scale features. Then the deep-features are sent to U-transformer for feature-level reconstruction. Finally and the reconstruction error is used to obtain the anomaly score map.

Source: [3] Chen, Liyang, et al. "Utrad: Anomaly detection and localization with u-transformer." Neural Networks 147 (2022): 53-62.

Introduction

In this study, we conduct experiments on real industrial datasets.

Our experiments focus on:

1. Looking for promising **machine learning methods for anomaly detection**
2. Test some **XAI models for anomaly interpretation**.

Methods

Introduction

In this study, we take a case from an industry system.

This system produces the fuel tank.

The image describes the thermal of the opening cap of the tank.

1. The dataset has two sets
   - Good set: 862 normal images.
   - Real set: 15453 normal + abnormal images
2. Each image has the same shape (256, 320)

an image in the good set.

an image in the real set.

Methods

Problem description:

In this problem, we have to complete two tasks:

1. Detecting abnormal image
2. Explaining the location of the anomaly in the image.

Methods:

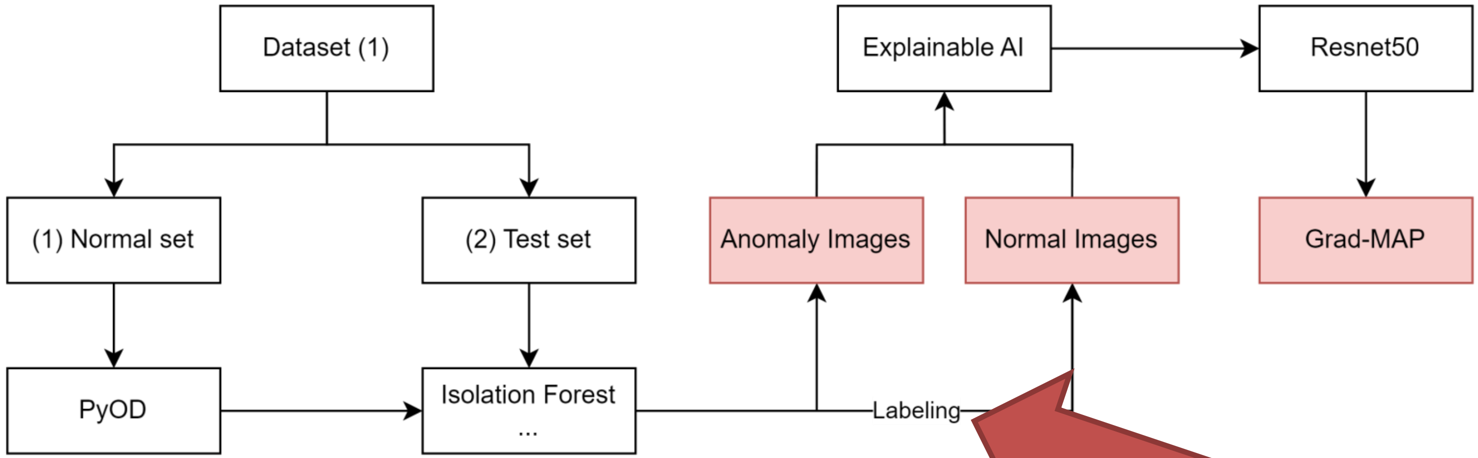
1. Two models: one for detecting, one for explaining
   - Detecting model: we use a toolbox (PyOD [1]) to run the 15 latest machine learning models
   - Explaining model: After we label the image based on the results of the detecting model. We use two types of Explainable AI models for explaining: LIME [2] and Grad-CAM[3]

2. Everything is one (EIO): detecting and explaining
- We use the U-transformer model (UTRAD: Anomaly detection and localization with U-Transformer)

[1] PyOD: A Python Toolbox for Scalable Outlier Detection

Method 1: Two model in pipeline

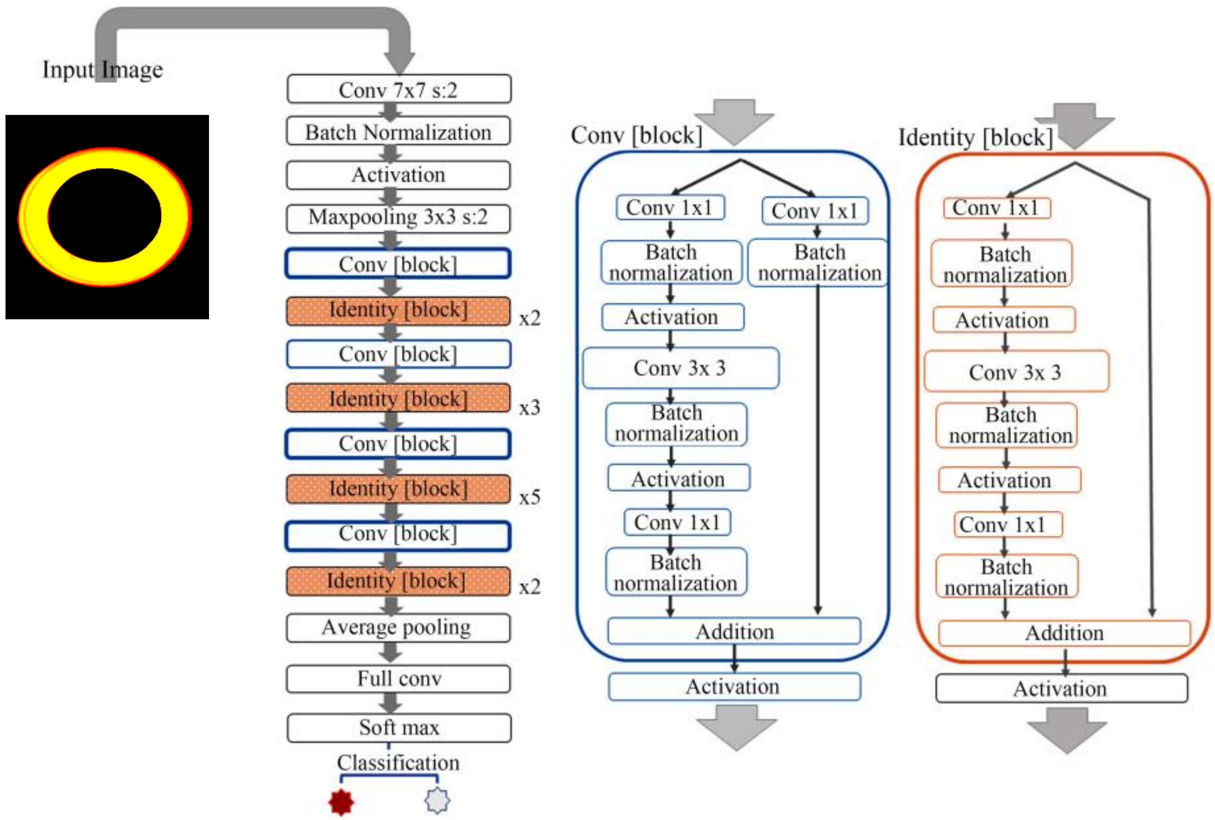
Two model frame work



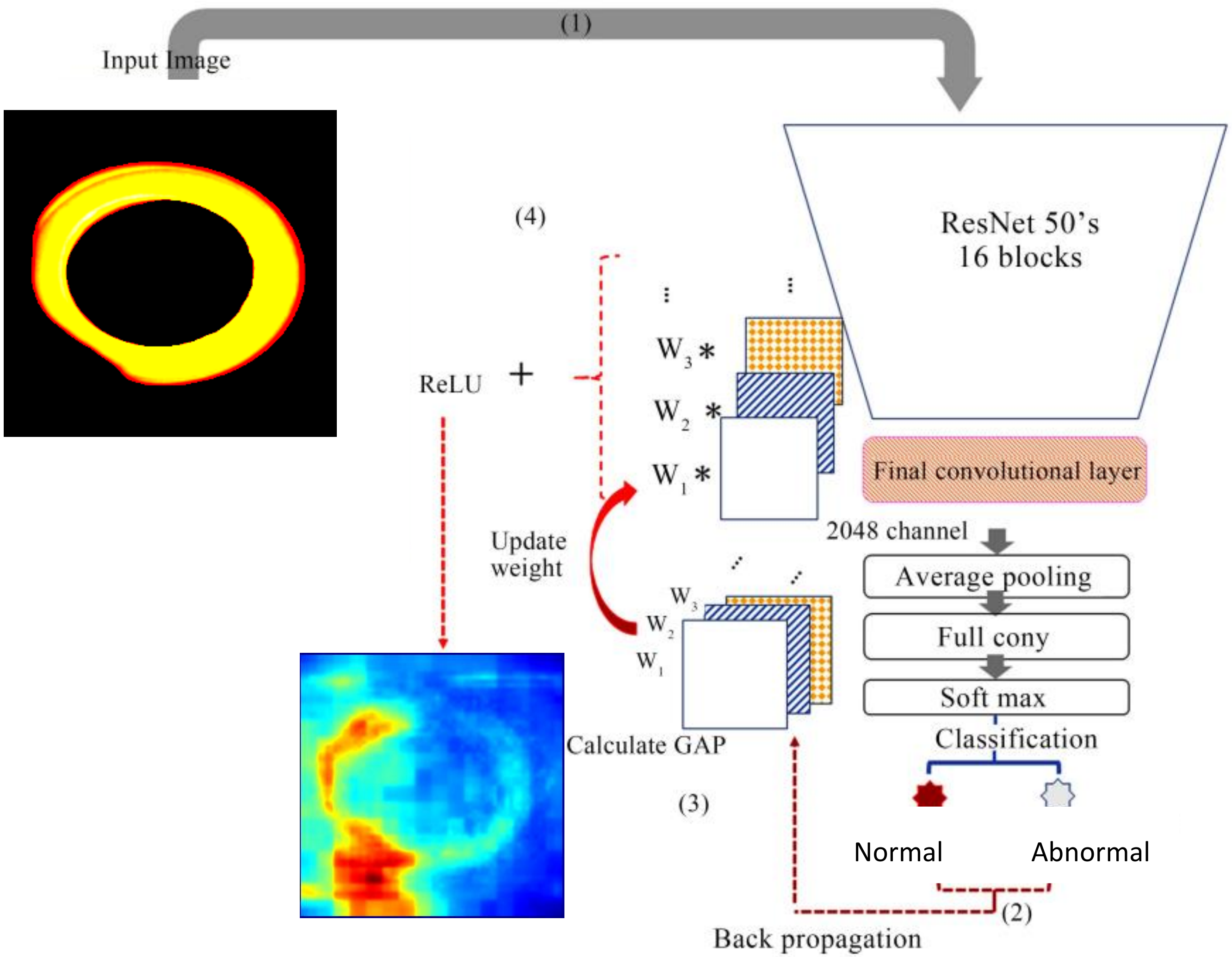
Detection models: Investigate **15** latest machine learning m s.

1. Angle-based Outlier Detector (ABOD)
2. Cluster-based Local Outlier Factor (CBLOF)
3. Feature bagging
4. Histogram-base Outlier Detection (HBOS)
5. Isolation Forest
6. K Nearest Neighbors (KNN)
7. Average KNN
8. Local Outlier Factor (LOF)
9. One-class SVM (OCSVM)
10. Principal Component Analysis (PCA)
11. Locally Selective Combination (LSCP)
12. Isolation-based anomaly detection using nearest-neighbor ensembles (INNE)
13. Gaussian Mixture Model (GMM)
14. Kernel Density Estimation (KDE)
15. Linear Model Deviation-base outlier detection (LMDD)

Explaining models: ResNet50 for classification



ResNet50 + Grad-CAM



Method 2: All-in-one model

The optimal separable two model does not ensure global optimization

⇒ Using All-in-one model

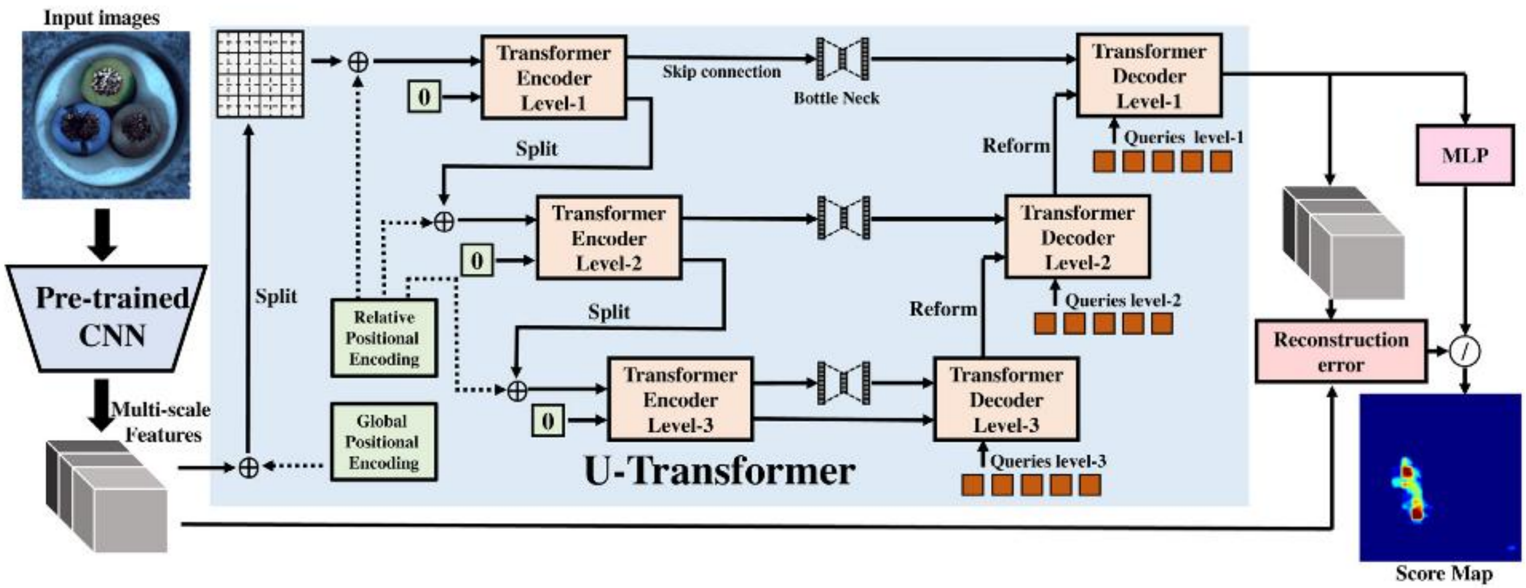


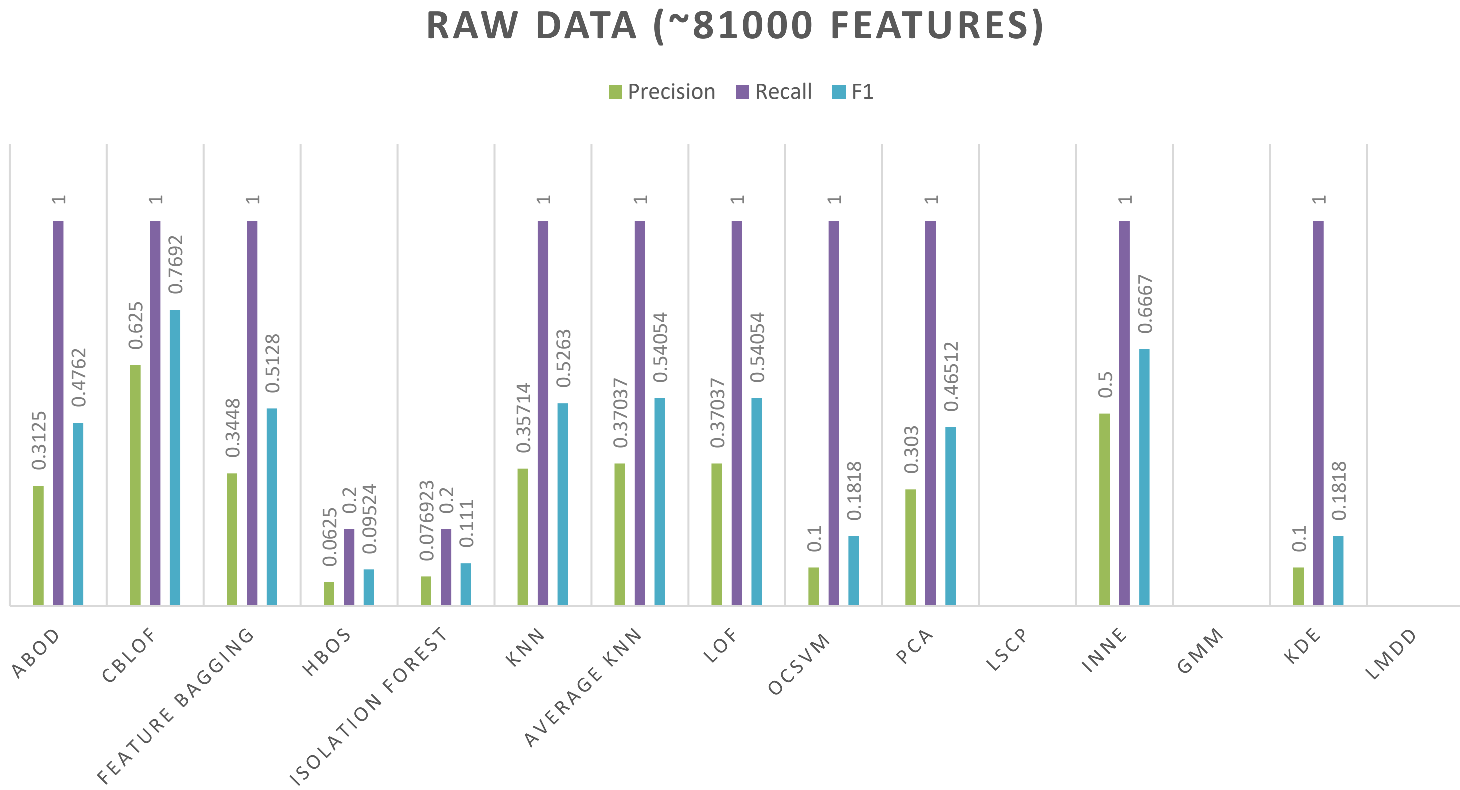
Fig. 1. The overall framework of the proposed anomaly detection and localization method. Firstly, The image is input to the pre-trained network to extract multi-scale features. Then the deep-features are sent to U-transformer for feature-level reconstruction. Finally and the reconstruction error is used to obtain the anomaly score map.

Results

Method 1: Performance of detection models

Case 1: Raw input (~81000 features)

Converting 2D images (256x320) to 1D vector (81920)

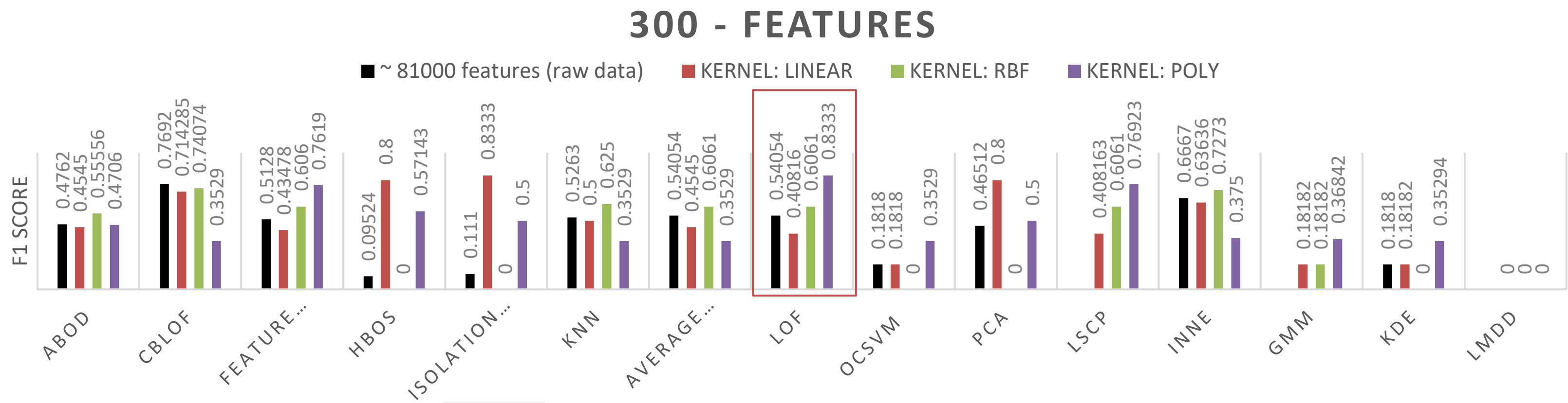


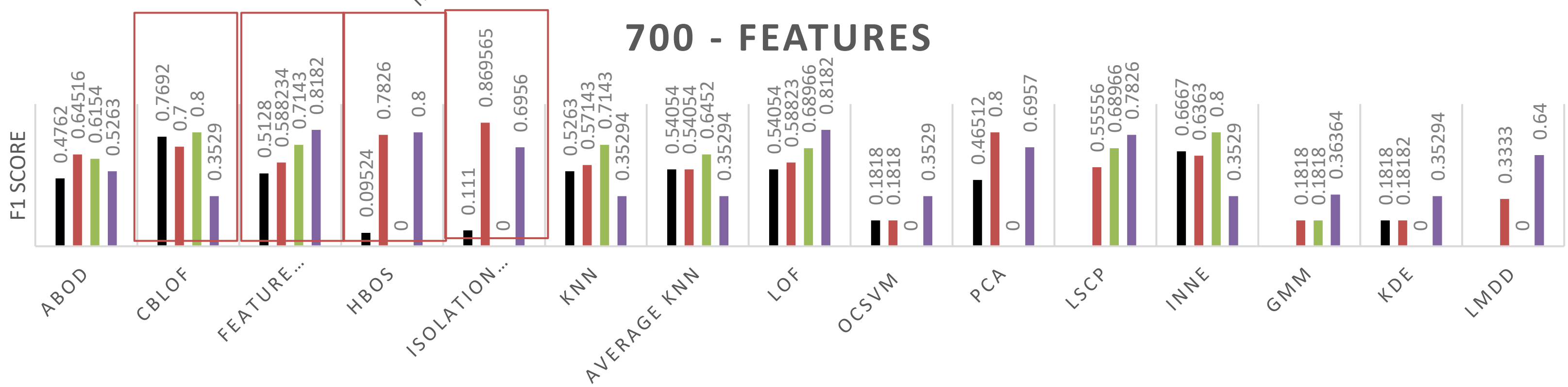
The performance of 15 methods with raw data

Kernel compare

Apply data preprocessing techniques:

1. Using **PCA** for reducing dimensions of data.
2. Mapping the output of PCA to another space based on **Kernel Function**





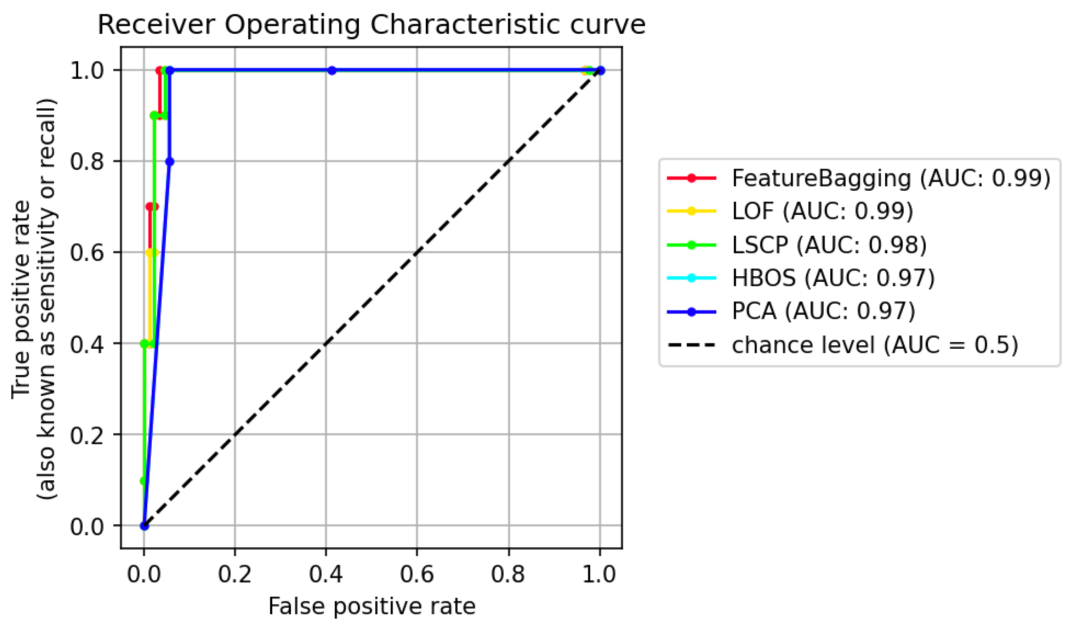
Performance Detection model

Top 5 methods (based on F1 scores – threshold=0.5):

1. Isolated forest
2. Local Outlier Factor (LOF)
3. Feature bagging
4. Cluster-based Local Outlier Factor (CBLOF)
5. Histogram-base Outlier Detection (HBOS)

Top 5 methods (based on AUC scores – 700 features):

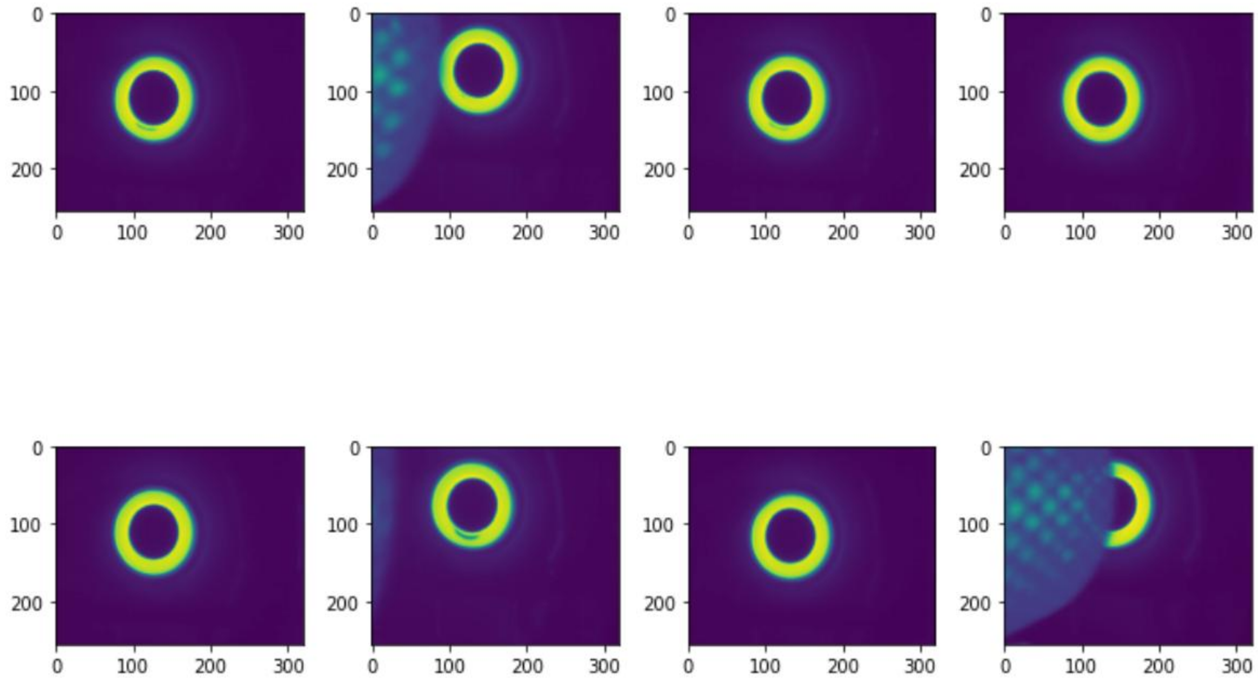
Comparing 5 machine learning methods in test set labeled by ourselves



Detection model

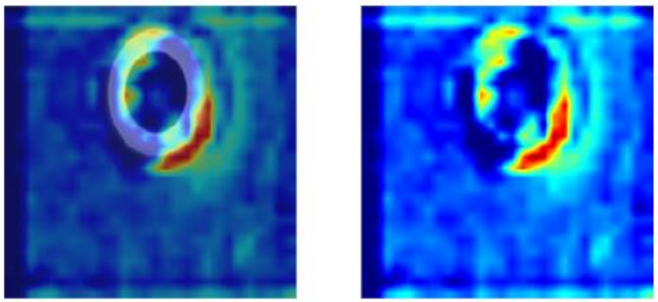
The result when applied Isolated forest (IF) in the real set

IF detects **5 abnormal** images in the **first 1000 images** of the real set.

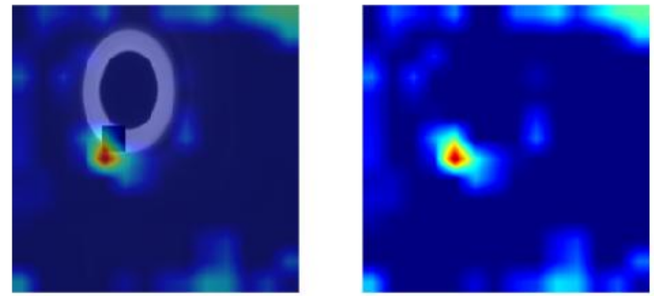


Explanation (Grad - CAM)

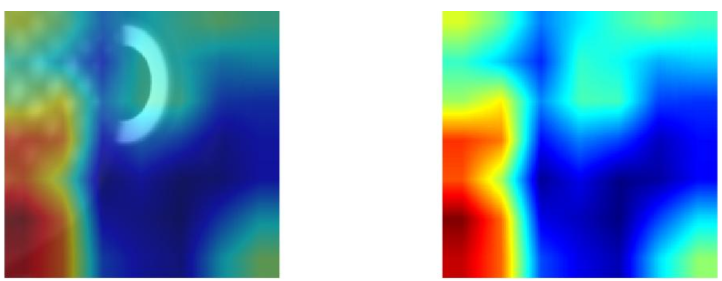
Normal



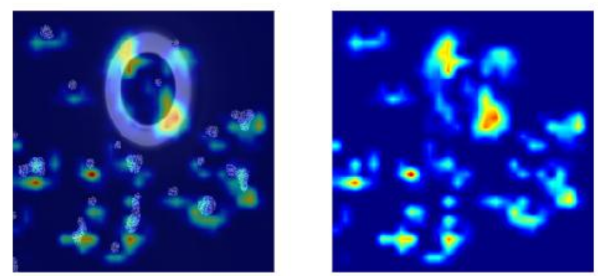
In the case of the object's shape being abnormal.



In the case of something strange caught on camera.



In the case of the image being noisy.

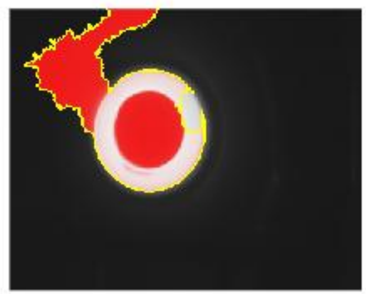


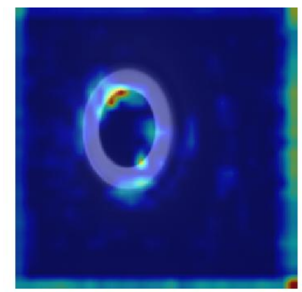
In the case of the object's location changing.

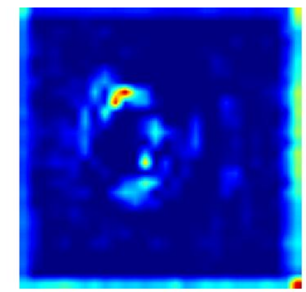
LIME

Grad-CAM

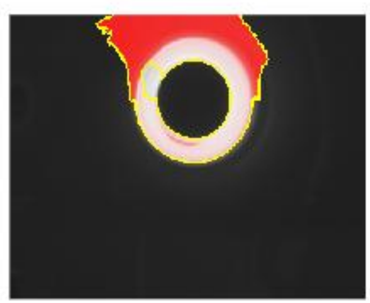
Actual: Normal
Predict: Anomaly

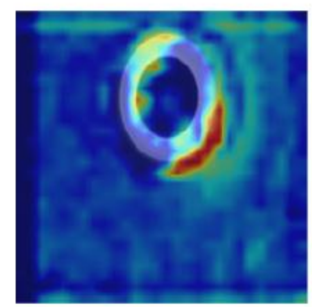


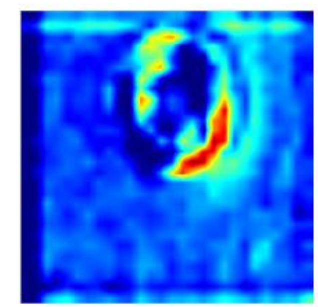




Actual: Normal
Predict: Normal

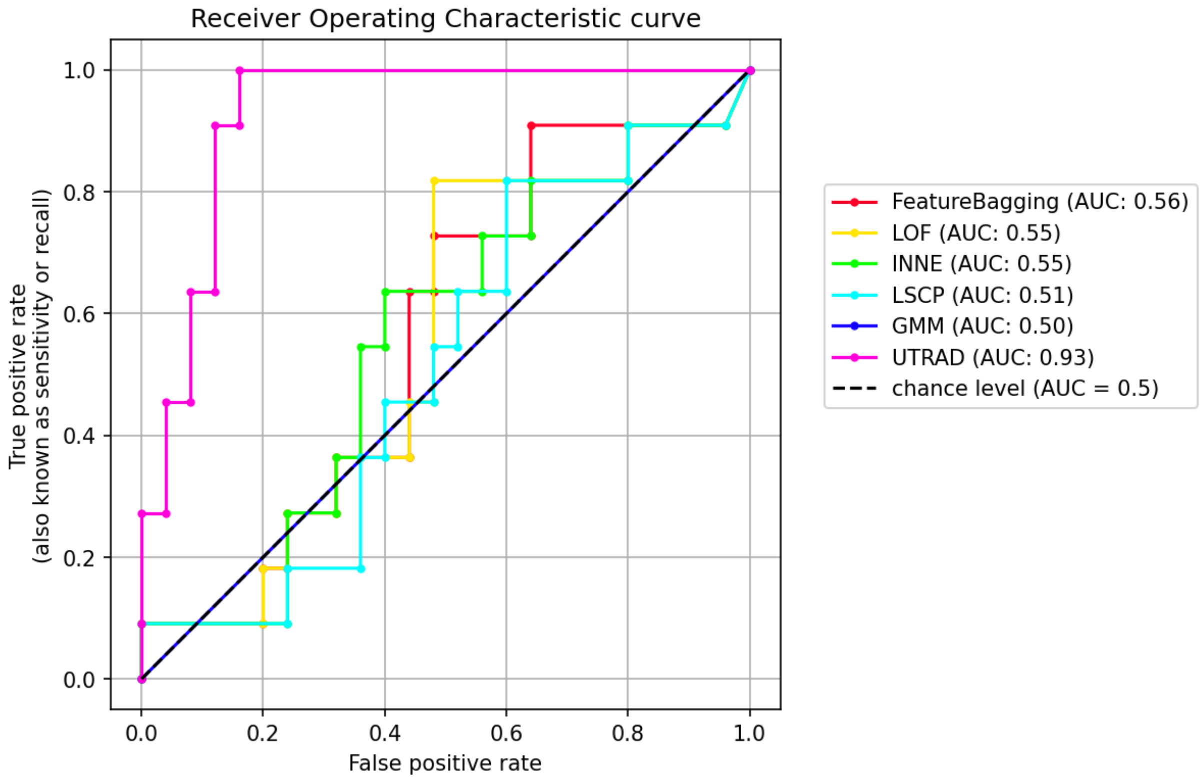






Result 2: AIO model

Comparing 5 machine learning methods vs UTRAD
in data **labeled by the expert**



Thank for listening