
Università degli Studi di Padova – Dipartimento di Ingegneria dell'Informazione
Corso di Laurea in Ingegneria dell'Automazione e dei Sistemi

Relazione per la prova finale
***«Unsupervised Anomaly Detection Approaches in
Computer Vision»***

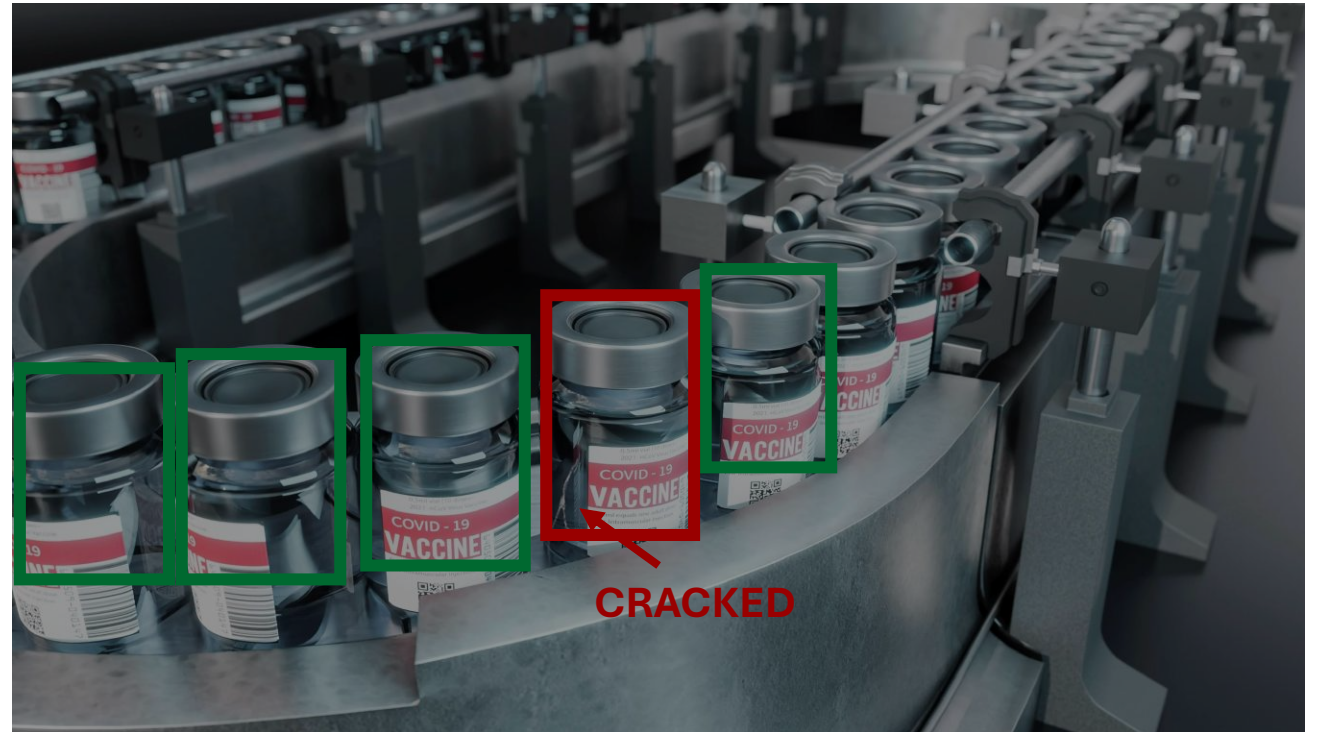
Relatore: Prof. Gian Antonio Susto
Correlatore: Dott. Davide Dalle Pezze

Laureando: Gioele Bonato
Matricola n.: 2101548

Padova, 20/07/2026

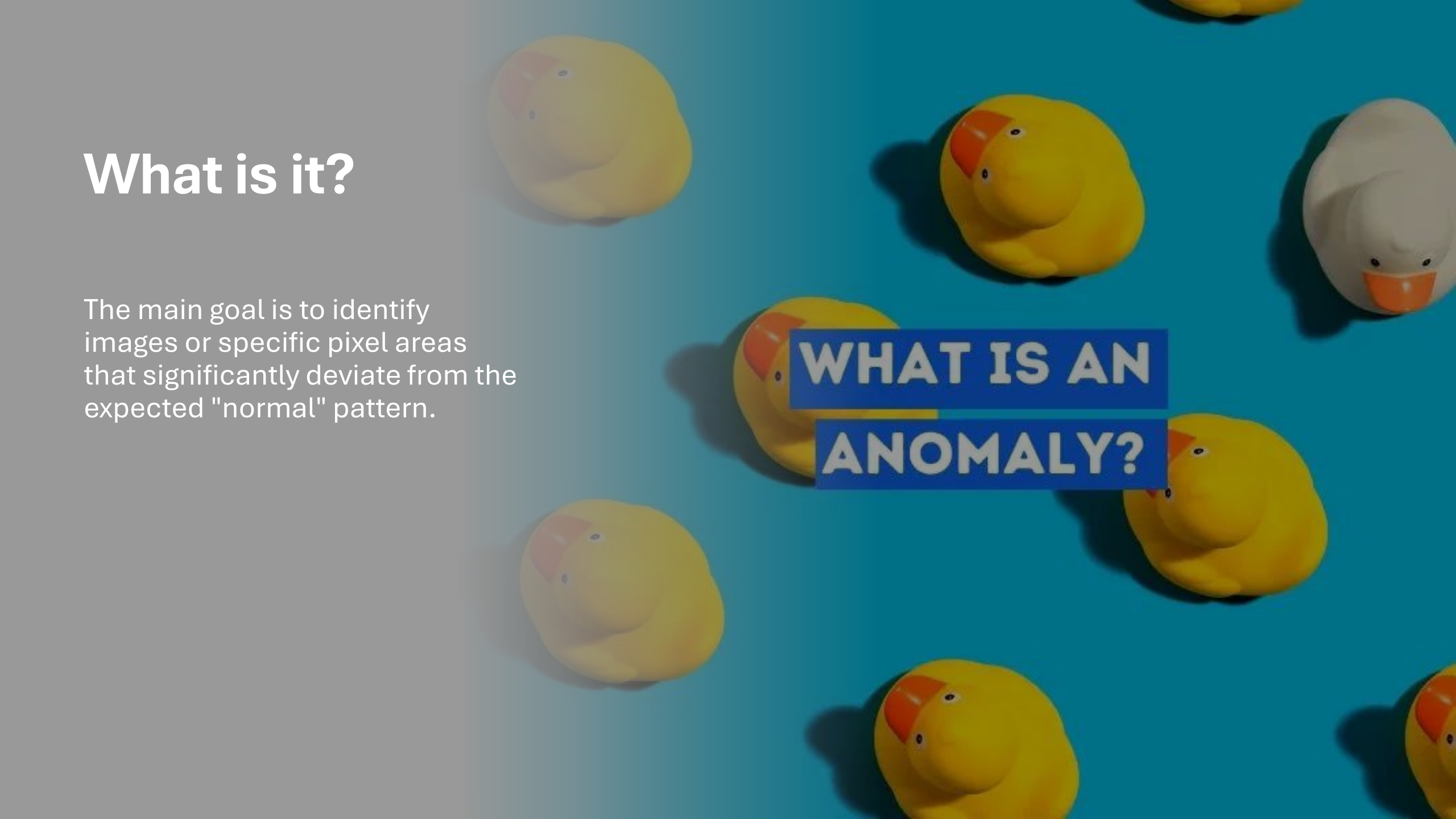
Topics

- Definition
- Applications
- Anomaly type
- Approches
- Datasets
- Evaluation Metrics
- Models
- Future Horizons



What is it?

The main goal is to identify images or specific pixel areas that significantly deviate from the expected "normal" pattern.

The background of the slide is a solid blue color. Scattered across this background are several yellow rubber ducks. In the upper right corner, there is one white rubber duck with an orange beak, which stands out from the rest of the yellow ducks. In the center of the slide, there is a blue rectangular box containing the text "WHAT IS AN ANOMALY?" in white, bold, uppercase letters. The box is split into two lines: "WHAT IS AN" on the top line and "ANOMALY?" on the bottom line.

**WHAT IS AN
ANOMALY?**

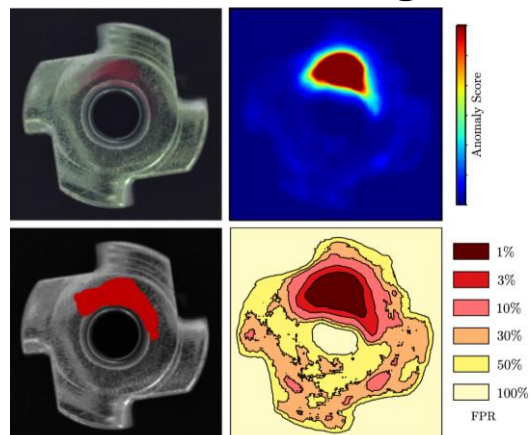
Why is it important?

In real-world scenarios, normal data is abundant while anomalies are rare, unpredictable, and can take infinite shapes (**Unsupervised enviroment**).

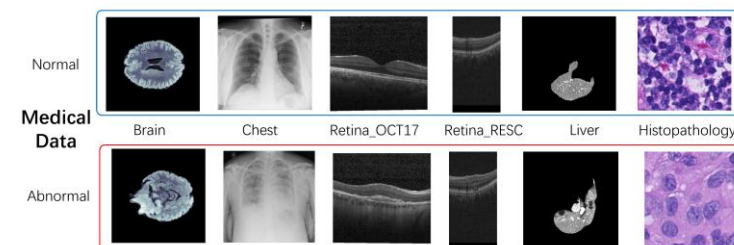
water leak 99.97%



Manufacturing

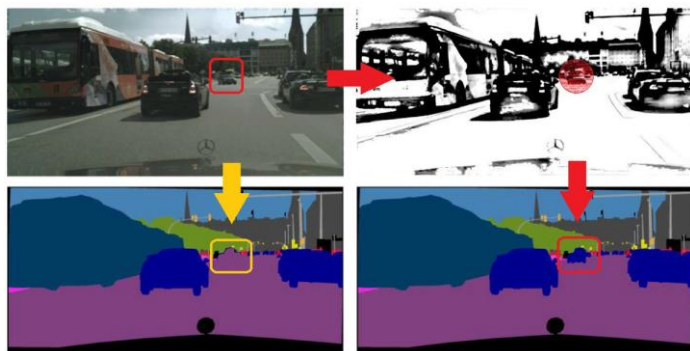


Medical

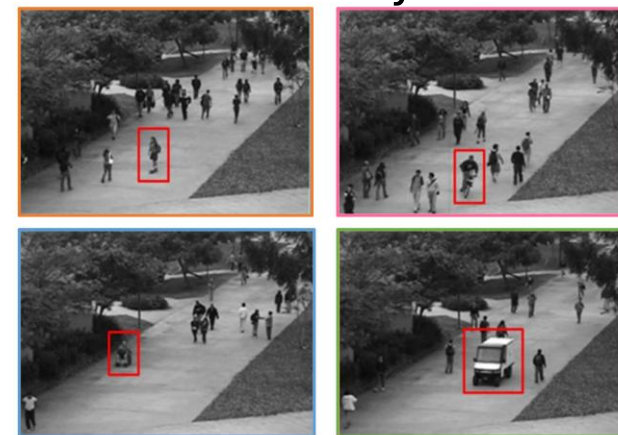


Applications

Automotive

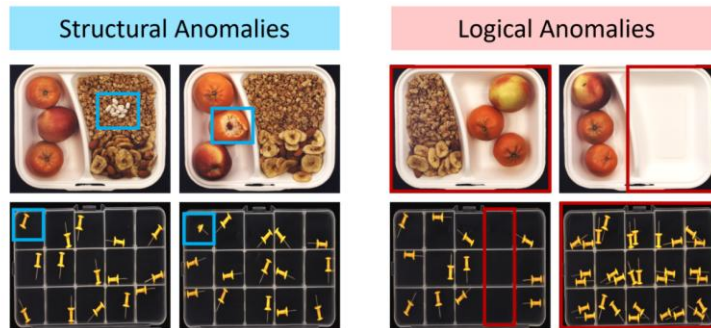


Security



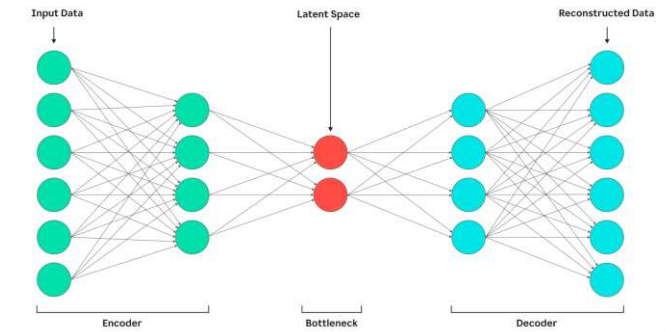
How to classify a method

Anomaly Type:

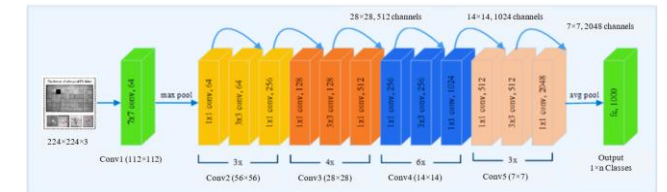


Techniques:

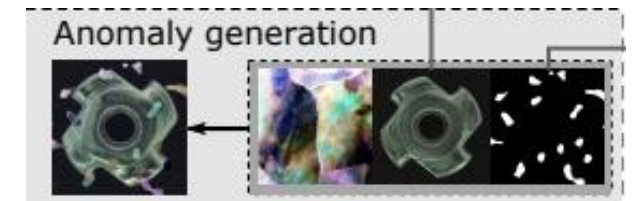
Reconstruction-based



Embedded-based



Synthesising-based



Datasets

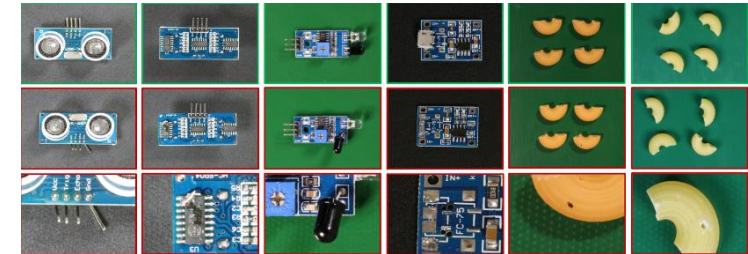
MVTec (2D) 2019

- 5.354 images (3.629 train, 1.725 test)
- 15 categories (10 objects, 5 texture)
- 73 defects



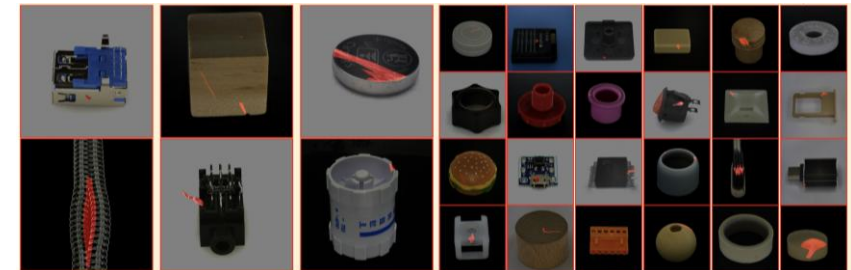
VisA 2022

- 10.821 images (9.621 normal, 1.200 anomalous)
- 12 categories (3 domains)
- 5-20 image per defect type



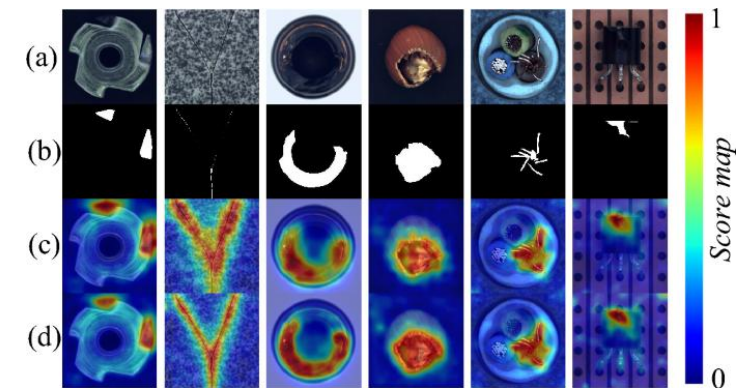
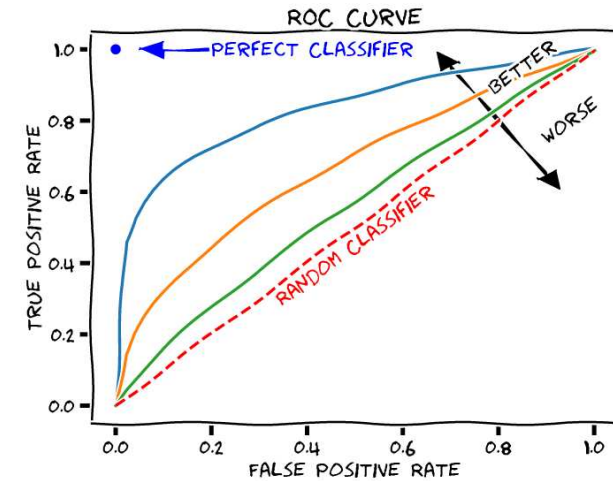
Real-IAD 2024

- 151.050 images (99.721 train, 51.329 test)
- 30 categories
- ~ 4.4 defects per category
- 5 shooting angles



Evaluation Metrics 1/2

- The metrics can be pixel-level or image-level based.
- ROC AUC : These metrics provide a comprehensive assessment of an algorithm's ability to discriminate between normal and anomalous instances.
- PRO-score : evaluation metric that weights ground-truth regions of different size equally.



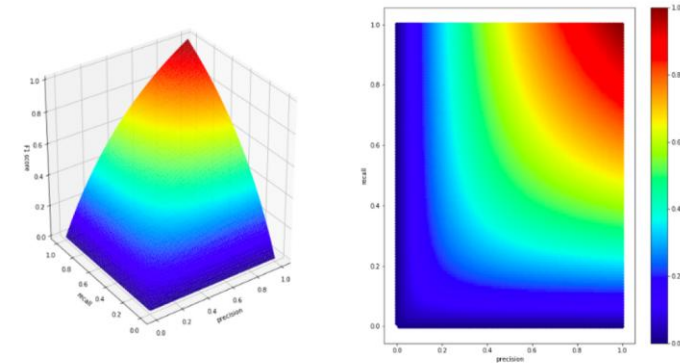
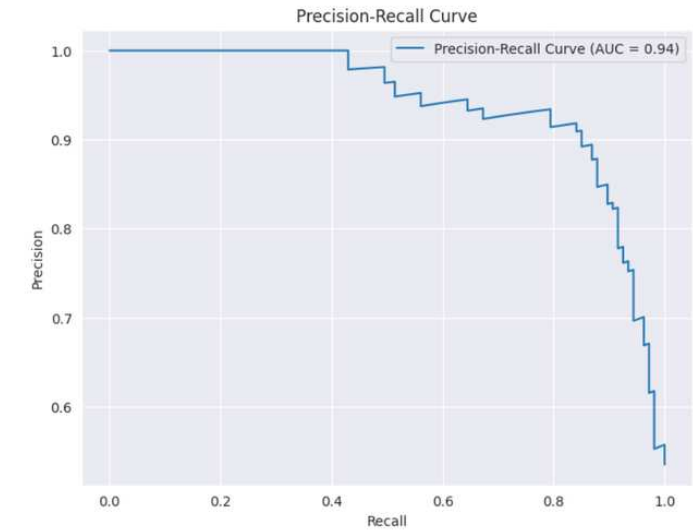
Evaluation Metrics 2/2

- Precision-Recall:

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN}$$

- F1-score : reflects both the model's ability to make accurate positive predictions and its ability to capture all relevant positive instances.

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall}$$

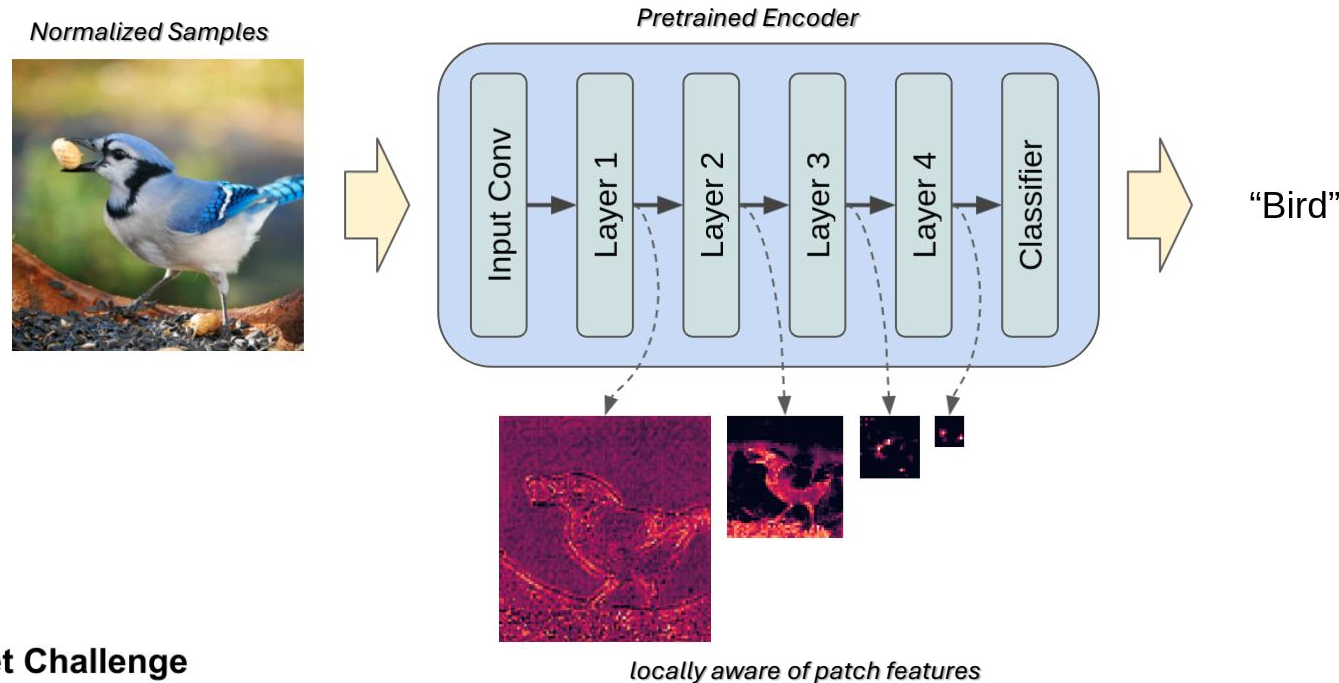


Model : PatchCore 1/3

1) Feature Extraction at Patch Level :

This approach consists of using a pre-trained network that was trained on ImageNet to extract the patch-level features from normal images in our dataset.

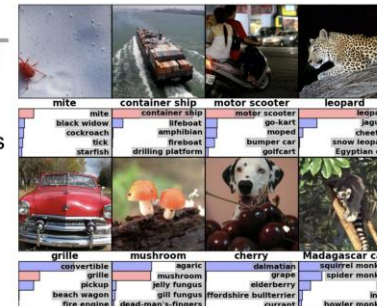
The patch-features are store in a Memory Bank.



ImageNet Challenge

IMAGENET

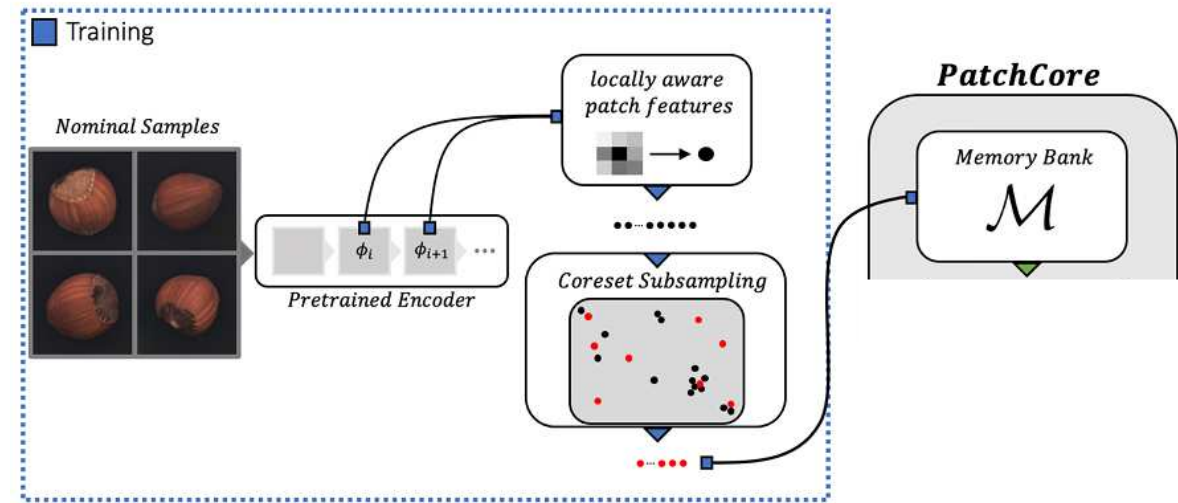
- 1,000 object classes (categories).
- Images:
 - 1.2 M train
 - 100k test.



Model : PatchCore 2/3

2) Coreset-reduction

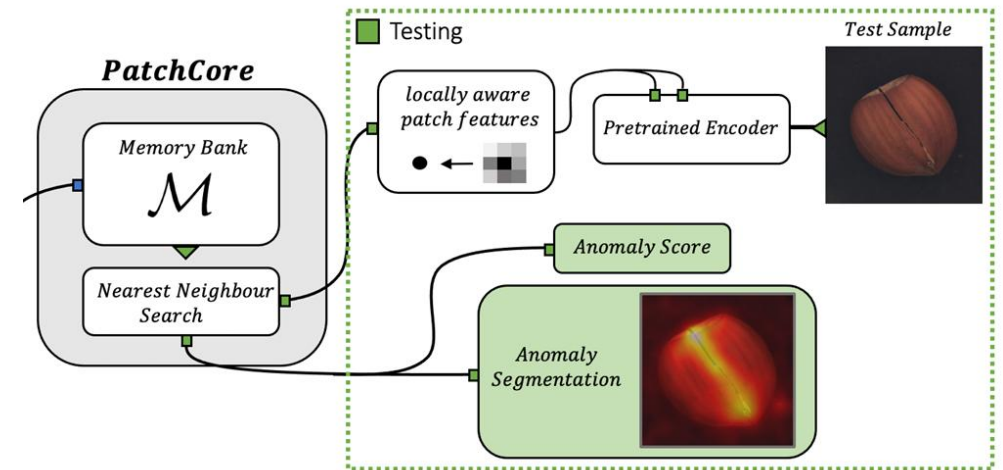
Large training datasets cause the Memory Bank to expand significantly, leading to high computational latency during inference. To accelerate the model, we reduce the Memory Bank size.



Model : PatchCore 3/3

3) Anomaly Detection (Inference)

- Feature Extraction : The pre-trained network extracts patch features from the test sample.
- Comparison : KNN compares each test patch with all elements in the Memory Bank.
- Anomaly Score : The distance to the nearest neighbor determines the local score (higher distance = greater anomaly).
- Decision : If any patch score exceeds a predefined threshold, the image is classified as anomalous.



Model : MambaAD 1/3

1) SSMs-based model

This types of models uses the math of control theory.

The SSMs-based model by their own has a low performance (too high accuracy). But the efficiency is good!!!

Mamba developer have found a strategy to make it work.

State equation

$$\mathbf{h}'(t) = \mathbf{A}\mathbf{h}(t) + \mathbf{B}\mathbf{x}(t)$$

Output equation

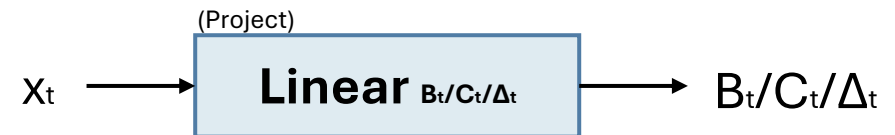
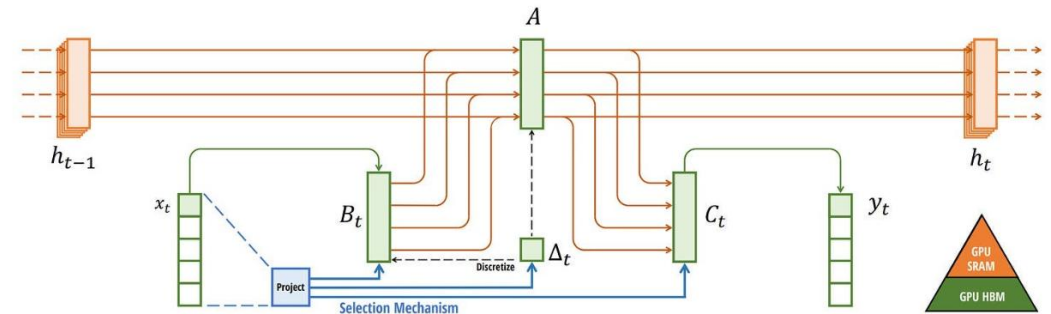
$$\mathbf{y}(t) = \mathbf{C}\mathbf{h}(t) + \mathbf{D}\mathbf{x}(t)$$

Model : MambaAD 2/3

2) Input Dependencies

In the way to have high performance is needed to use Selective SSMs, so a discretized and input dependent SSM.

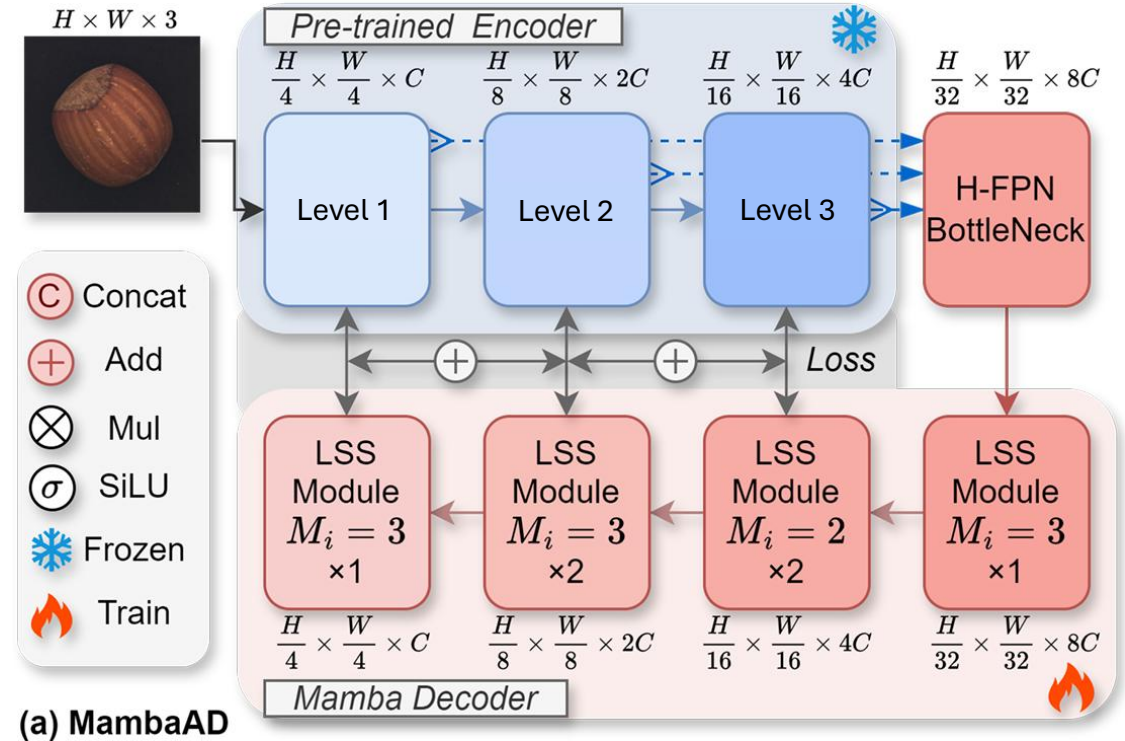
In this case we need to use the Parallel Associative Scan algorithm (fast on GPUs).



Model : MambaAD 3/3

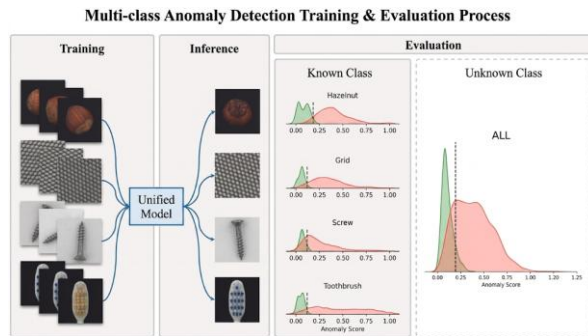
3) The Model

The model works using a Pre-trained Encoder from whom feature are extracted and organised in the BottleNeck, after we use the mamba block to reconstruct them.



Conclusions: Future Horizons

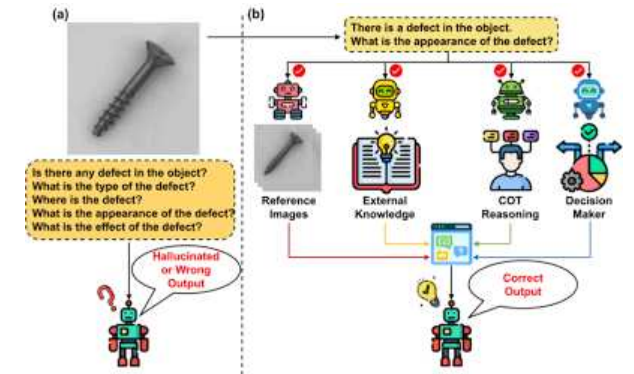
Unified multi-class framework



Computational efficiency - Edge Computing



Zero-Shot mode



Thank you for your attention

Special thanks to:

- Davide Dalle Pezze,
- Francesco Borsatti,
- Manuel Barusco
- Prof. Gian Antonio Susto

Extra Material 1.1

SSMs, further insights:

State Space models (control theory) are the mathematical tool on which Mamba bases its operation with some variations.

State equation $\mathbf{h}'(t) = \mathbf{A}\mathbf{h}(t) + \mathbf{B}\mathbf{x}(t)$

Output equation $\mathbf{y}(t) = \mathbf{C}\mathbf{h}(t) + \mathbf{D}\mathbf{x}(t)$

↓ Discretization

$$\overline{\mathbf{A}} = \exp(\Delta\mathbf{A})$$

$$\overline{\mathbf{B}} = (\Delta\mathbf{A})^{-1} (\exp(\Delta\mathbf{A}) - \mathbb{I}) \cdot \Delta\mathbf{B}$$

↓ Fast (Parallelizable)

$$\mathbf{K} = \left(\mathbf{C}\overline{\mathbf{B}}, \mathbf{C}\overline{\mathbf{A}}\overline{\mathbf{B}}, \dots, \mathbf{C}\overline{\mathbf{A}}^{t-1}\overline{\mathbf{B}} \right)$$

$$\mathbf{X} = (\vec{x}_1, \vec{x}_2, \dots, \vec{x}_L)$$

$$\mathbf{Y} = \mathbf{K} * \mathbf{X}$$

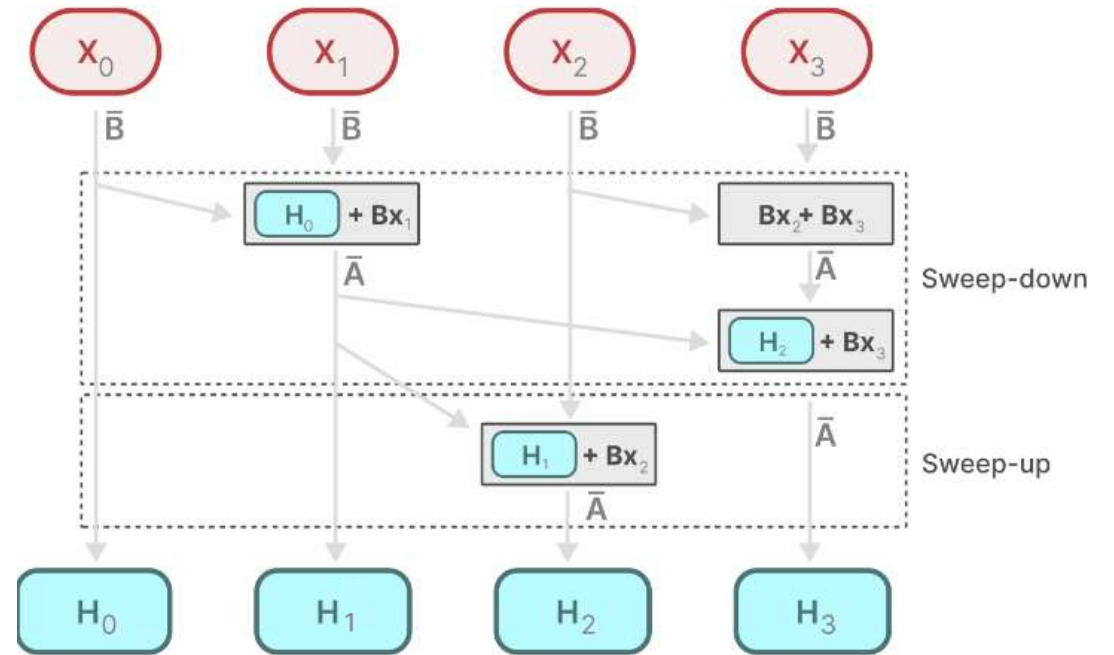
↓

**Efficient
but
Low performance**

Extra Material 1.2

From SSMs to Selective SSMs (Mamba core)

Parallel Associative Scan: Instead of processing patches step-by-step, Mamba uses a tree-structured approach (*Sweep-down* and *Sweep-up*) to compute hidden states (H_t) in parallel on the GPU.



Extra Material 2.1

MambaAD vs PatchCore

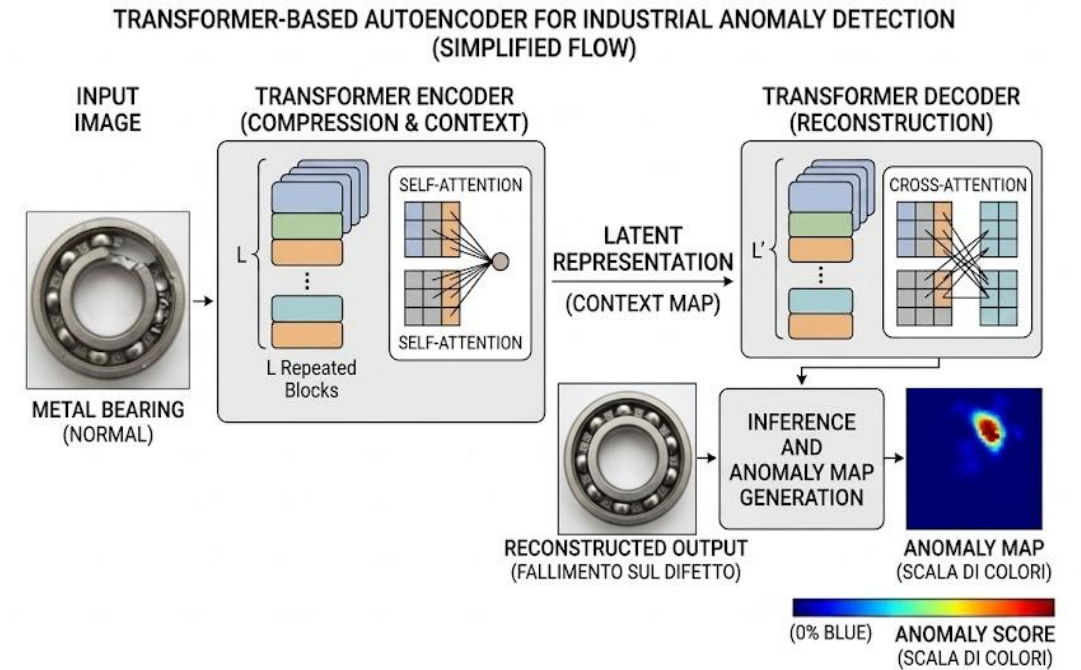
Tabella 1: Confronto prestazionale ed efficienza su benchmark MVTec AD.

Modello	AUROC	PRO-score	F1-score	Precision-Recall
PatchCore	99.1%	98.2%	98.0%	97.9%
MambaAD	98.5%	97.6%	97.2%	97.1%

Extra Material 3.1

Viual Transformers

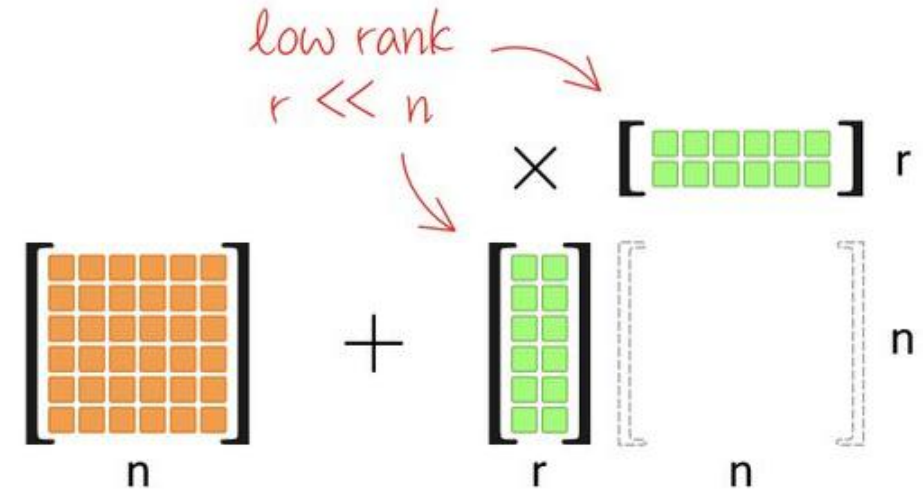
Transformers are reconstruction-based methods. This type of model represents the current state of the art for capturing long-range dependencies and whole-image context in anomaly detection.



Extra Material 4.1

LoRA

Efficient VLM Tuning: Instead of updating billions of parameters in a Vision-Language Model, LoRA freezes the original weights W_0 and trains only small, low-rank matrices A and B .



$$W = W_0 + \Delta W = W_0 + (B \times A)$$

Diagram illustrating the training status of the matrices in the equation $W = W_0 + \Delta W = W_0 + (B \times A)$. The term W_0 is labeled "frozen" with an arrow pointing to it. The term $B \times A$ is labeled "trainable" with two arrows pointing to B and A respectively.