University of Padova
Department of Mathematics “Tullio Levi-Civita”

MASTER THESIS
IN **COMPUTER SCIENCE**

Memory-Efficient Continual Learning for Visual Anomaly Detection: **A Compressed Replay Approach for Teacher-Student Architectures**

SUPERVISOR
Prof. Gian Antonio Susto

ASSISTANT ADVISORS
Dr. Davide Dalle Pezze
Ph.D. Student Manuel Barusco

CANDIDATE
Lorenzo D’Antoni

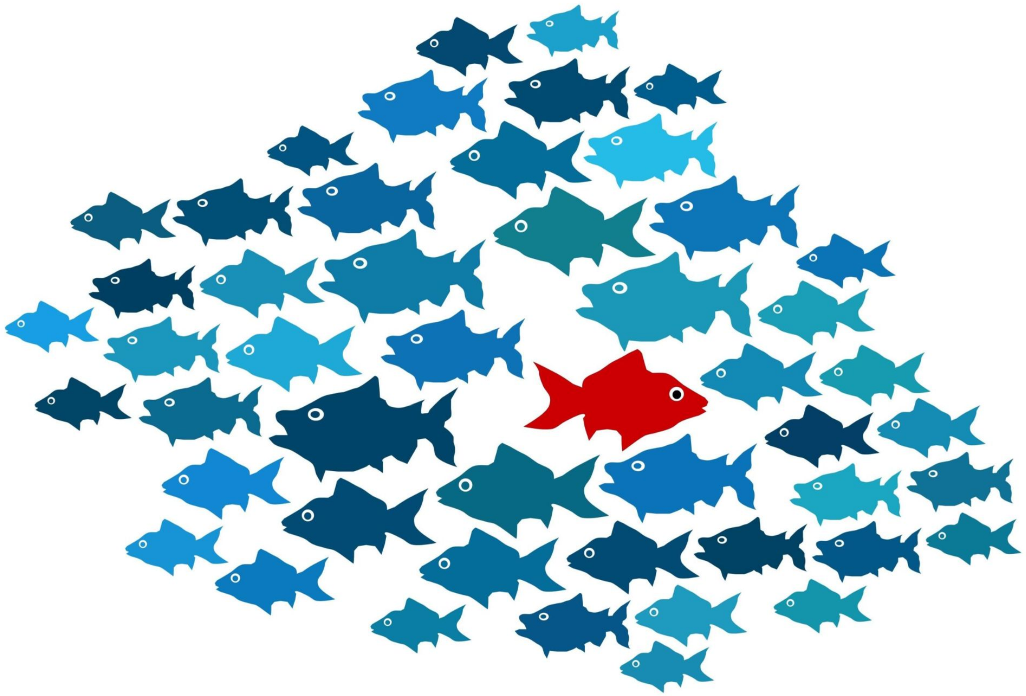
STUDENT ID
2073767

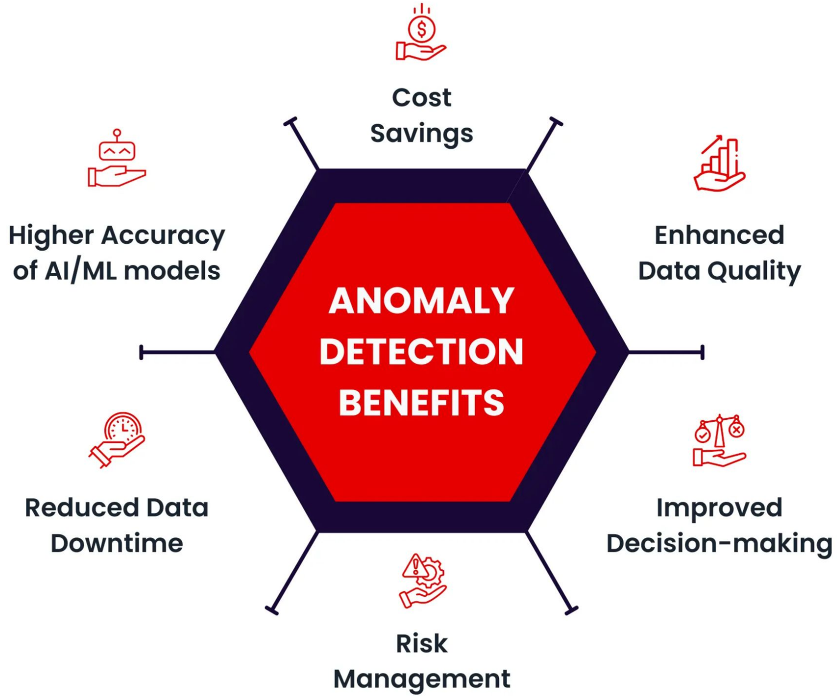
ACADEMIC YEAR
2023-2024

Anomaly Detection

Definition

Anomaly detection is the identification of **observations**, **events** or **data points** that deviate from what is usual, standard or expected



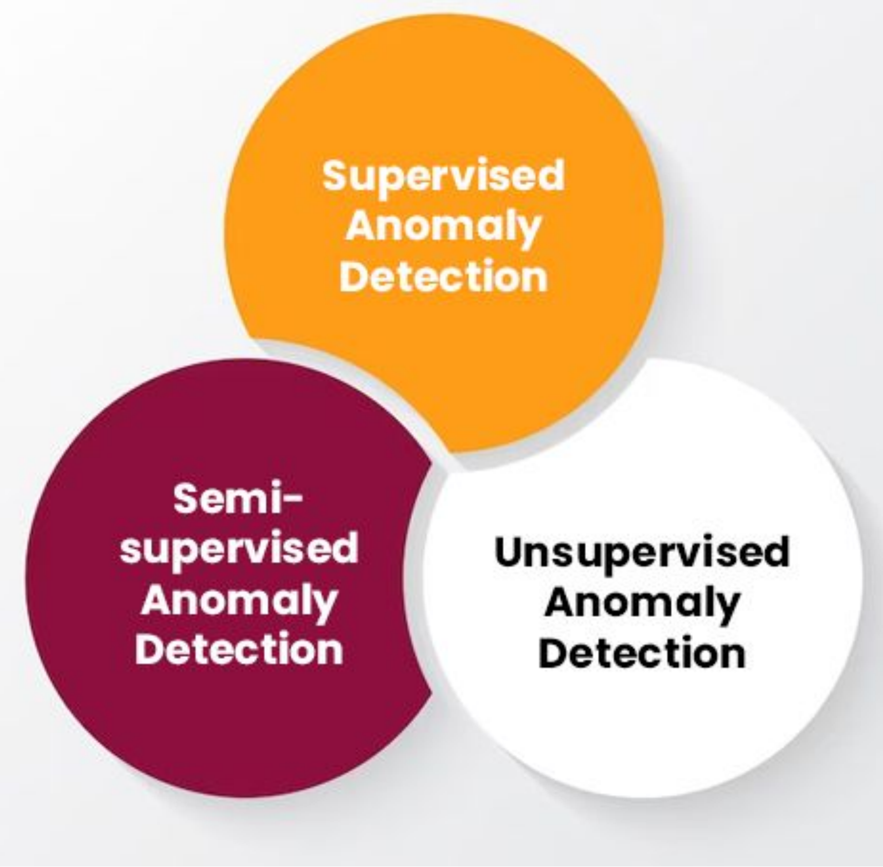


Anomaly Detection ML Strategies

Supervised: needs labeled data (normal + anomalous instances)

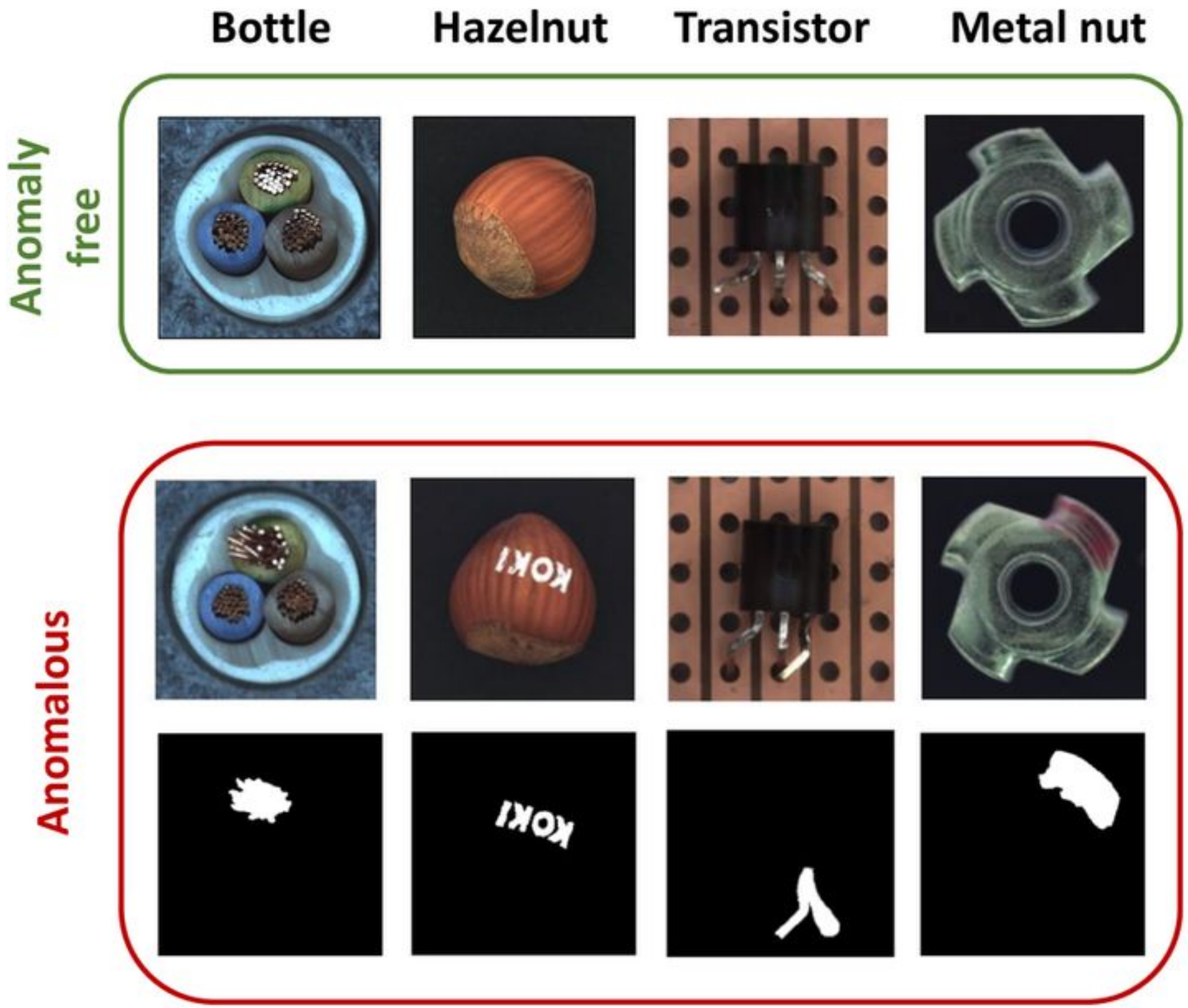
Semi-supervised: needs few labeled data (normal + few anomalous instances)

Unsupervised: unlabeled data (normal instances)



(Unsupervised) Visual Anomaly Detection

- **image-level**
  - anomalous or not?
  - not interpretable
- **pixel-level**
  - which pixels cause the anomaly?
  - interpretable

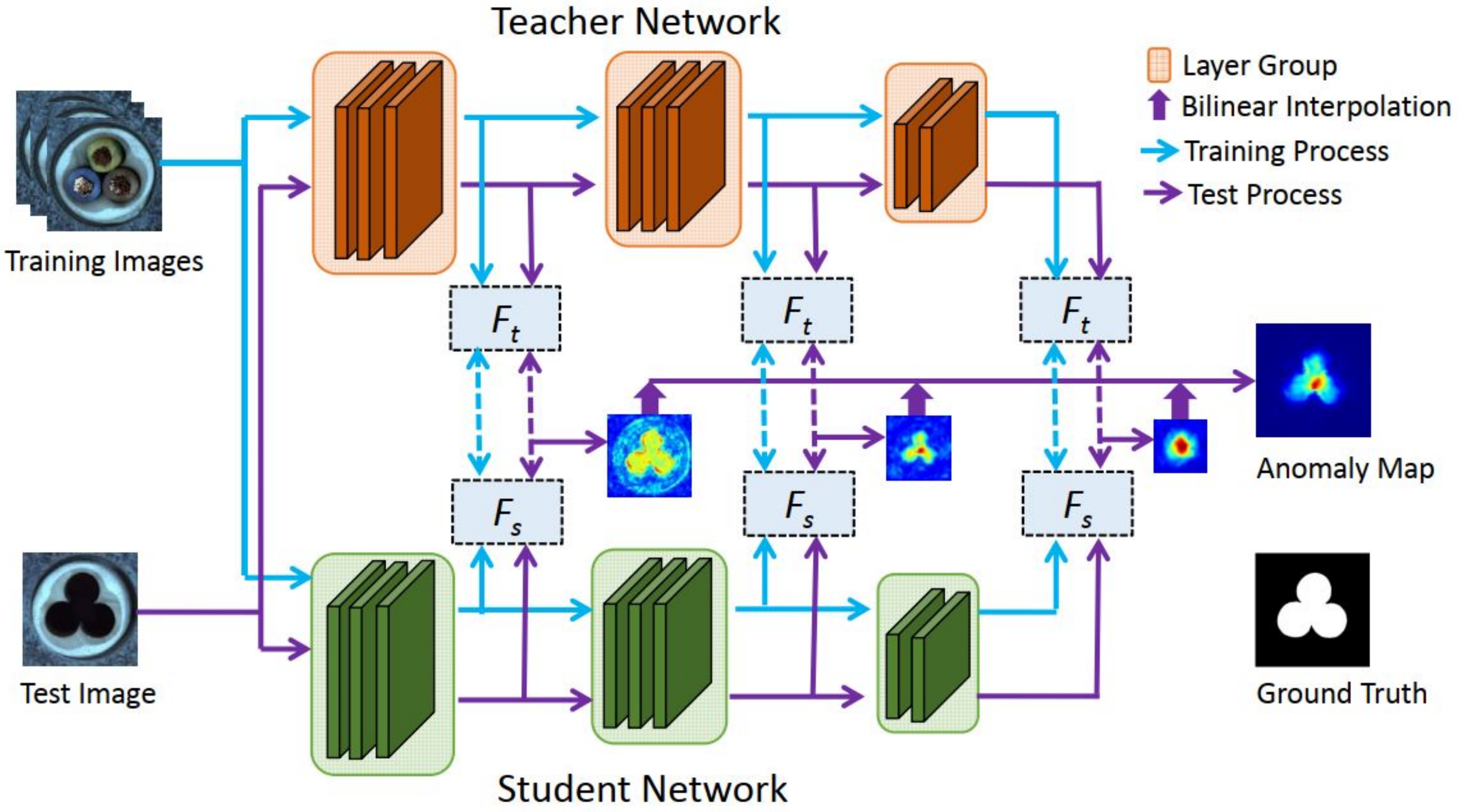


STFPM (Student-Teacher Feature Pyramid Matching)

UNIVERSITÀ DEGLI STUDI DI PADOVA

Feature-based distillation

replicates the teacher's feature representation on normal images.



First problem

Deep Learning is not a **one-size-fit-all**

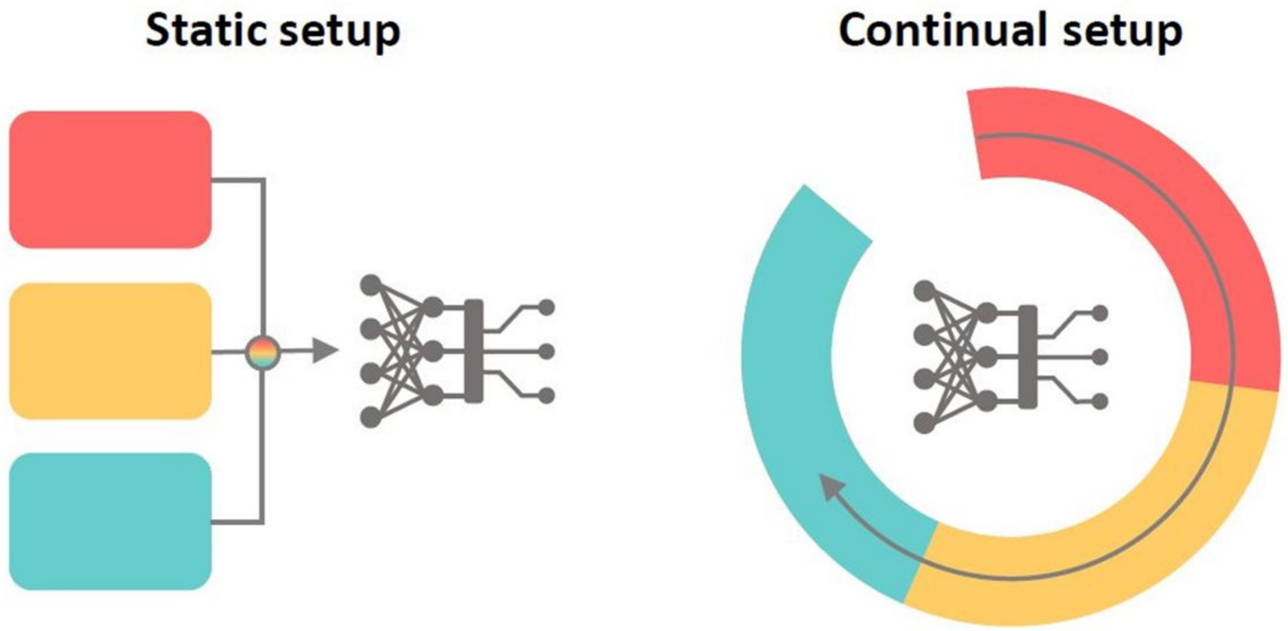
PROS	CONS
❑ handles high-dimensional, complex data	❑ computationally intensive
❑ captures spatial features and hierarchical relationships	❑ memory, inference time, power constraints limit performance
❑ automates feature extraction	❑ large datasets needed for training
	❑ limited interpretability

Second problem

Università degli Studi di Padova

Continual Learning strategies are needed since

- data distribution can shift
- new tasks may appear



Objective

Memory-efficient approach

for Continual Learning

in Visual Anomaly Detection

using Teacher-Student architectures

- reduce memory usage, speed up inference
- minimize catastrophic forgetting
- maintain high detection accuracy
- STFPM on MVTec AD dataset

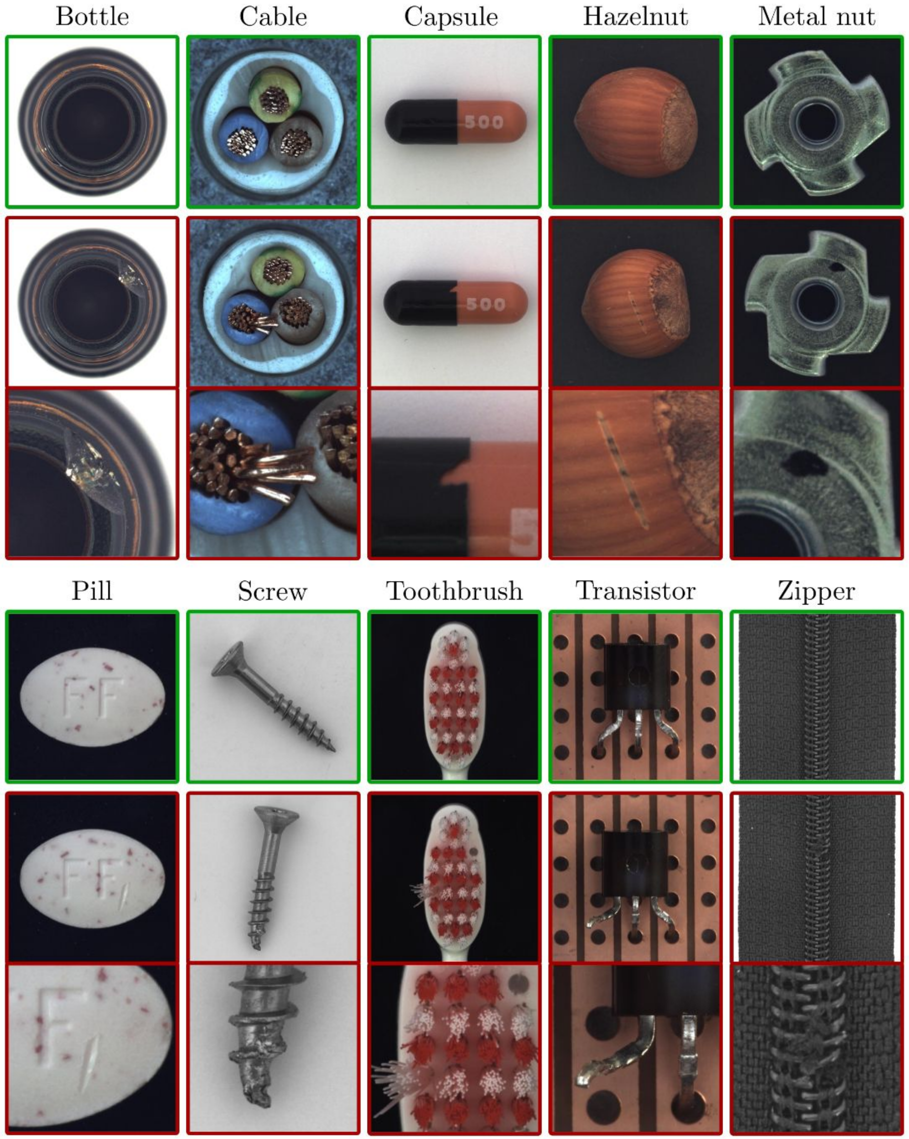
Exploratory Data Analysis

dataset:

- **MVTec Anomaly Detection**
- 10 objects
- scratches, dents, holes, stains, …

analyze features extracted by **STFPM**

gain valuable insights



Feature Importance

Compute the Importance of original features using **PCA** (Principal Component Analysis)

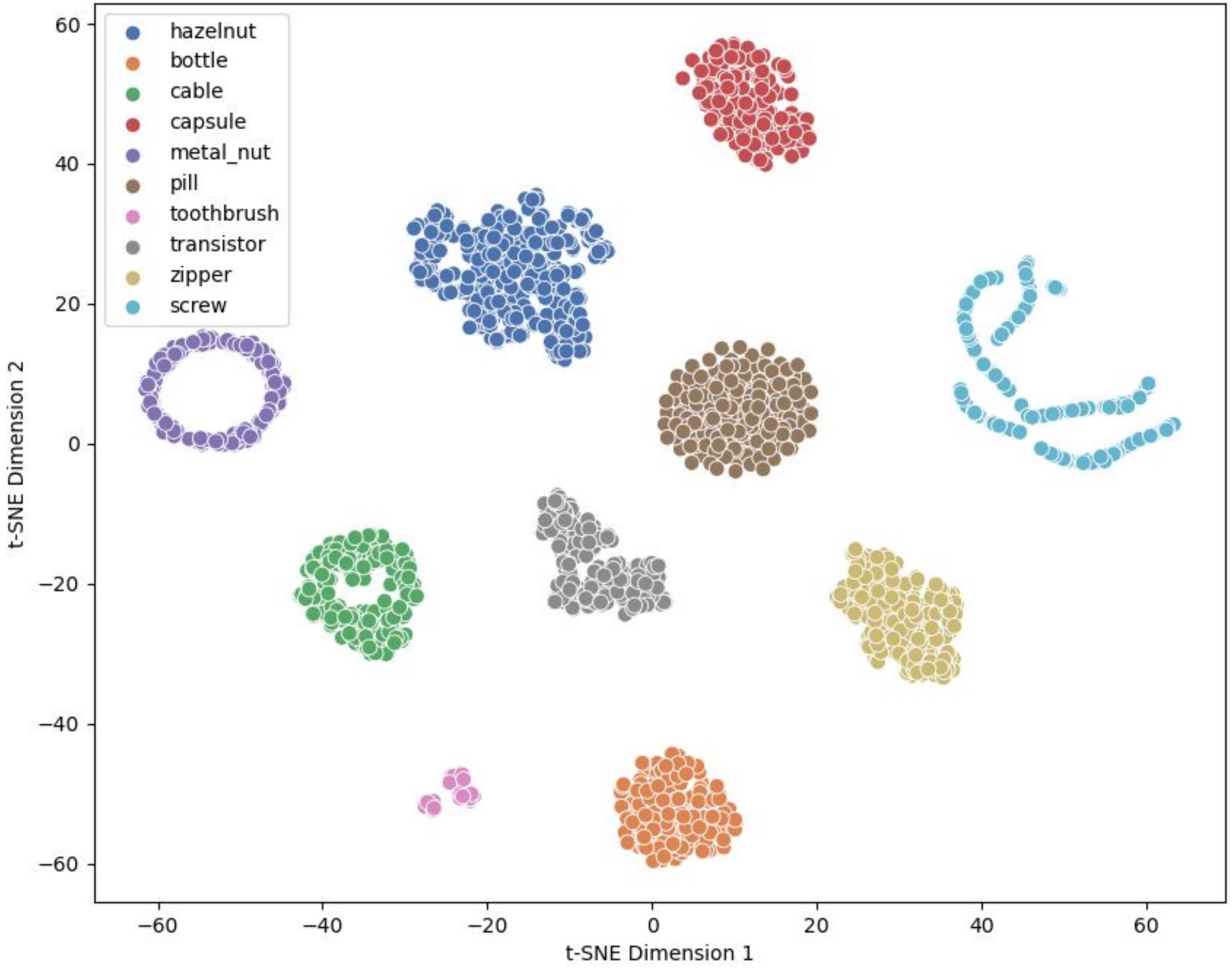
↓

Select only the **top p%** features

↓

Plot **t-SNE** graph (t-distributed Stochastic Neighbor Embedding)

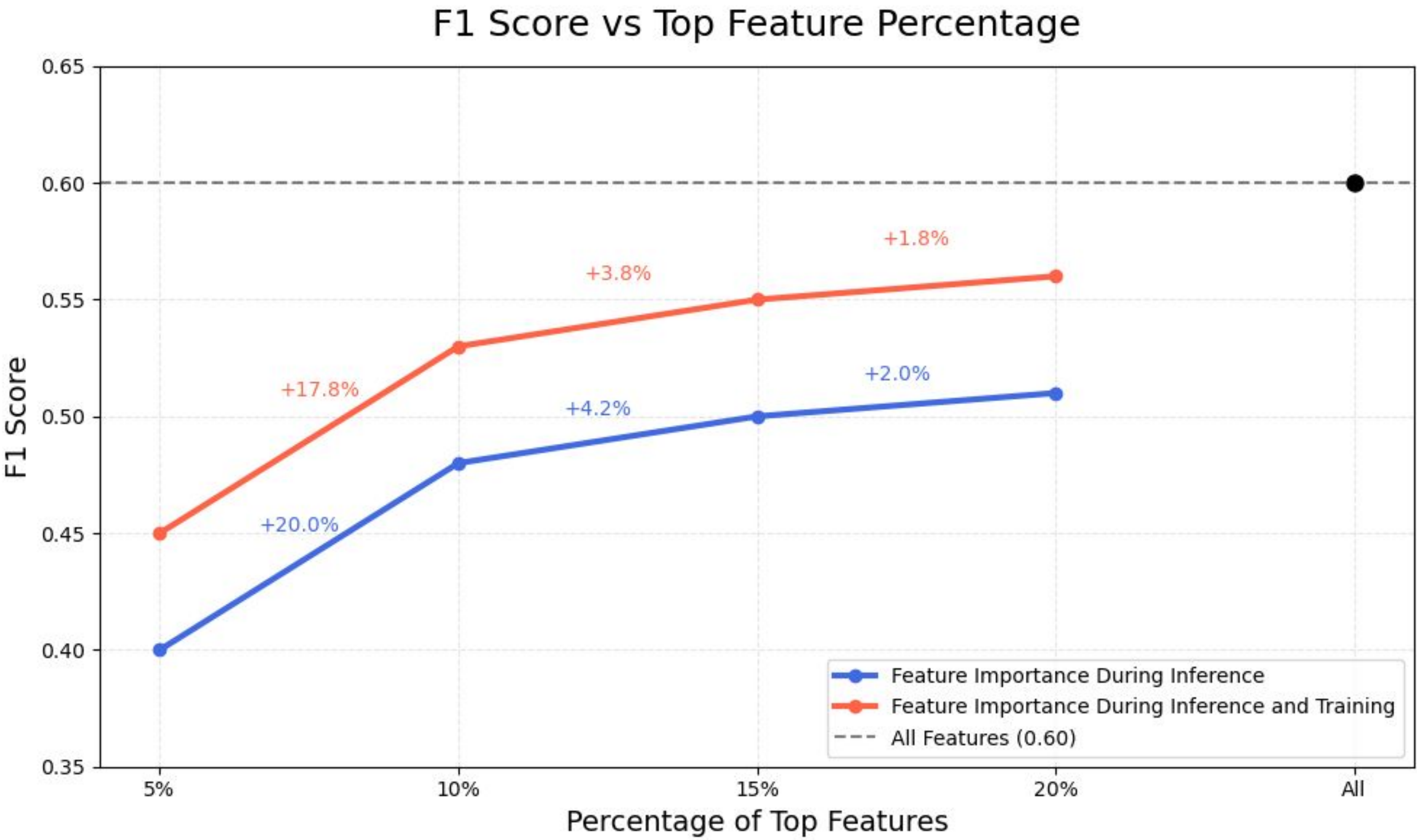
t-SNE plot (top 1% features)



Feature-Selective Fine-Tuning (STFPM)

Using **top-ranked features** preserves anomaly detection performance?

Objects provided sequentially (Naive Fine-Tuning)



Returning to the main objective

Università degli Studi di Padova

PaSTe (Partially Shared Teacher-Student)

MCUNet-in3 lightweight backbone network

AD literature (2024)

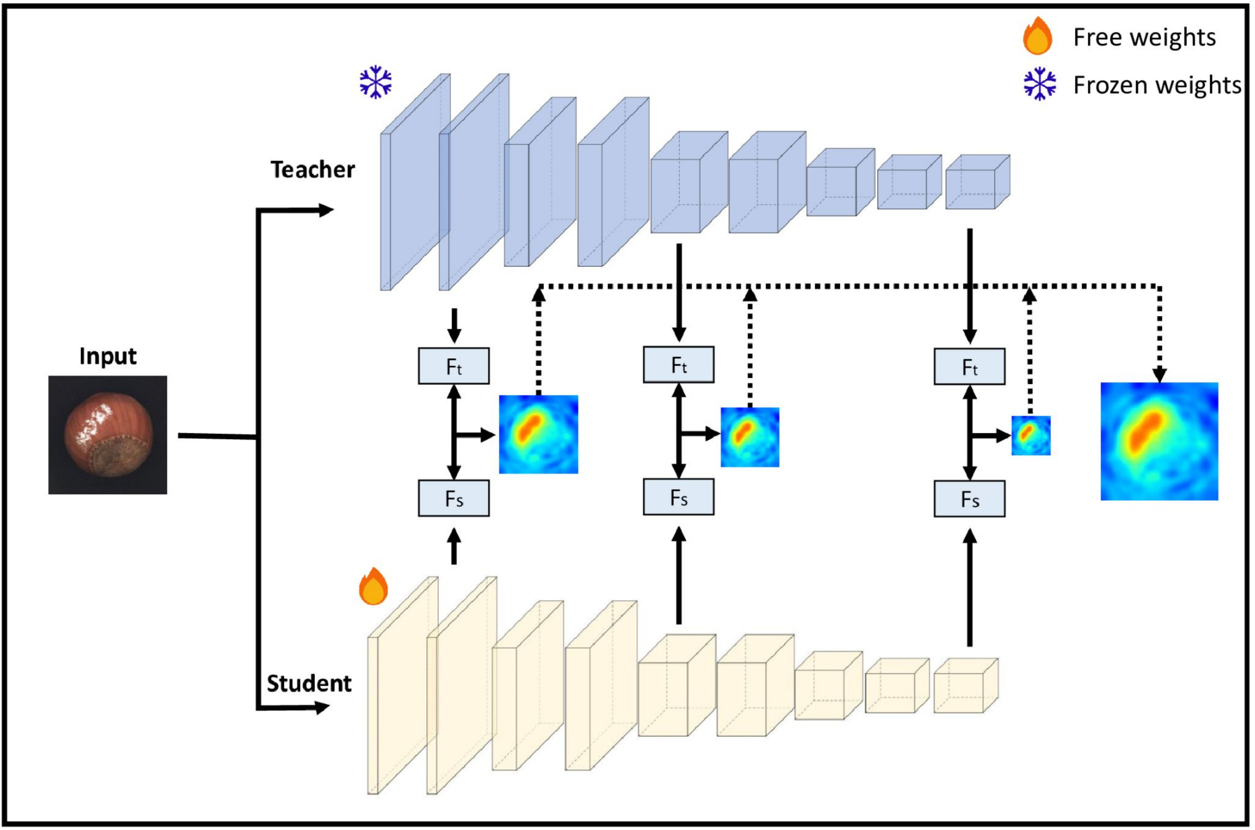
Adapt PaSTe to the **Continual Learning** scenario

Resource-efficient replay for edge devices (**compressed replay**)

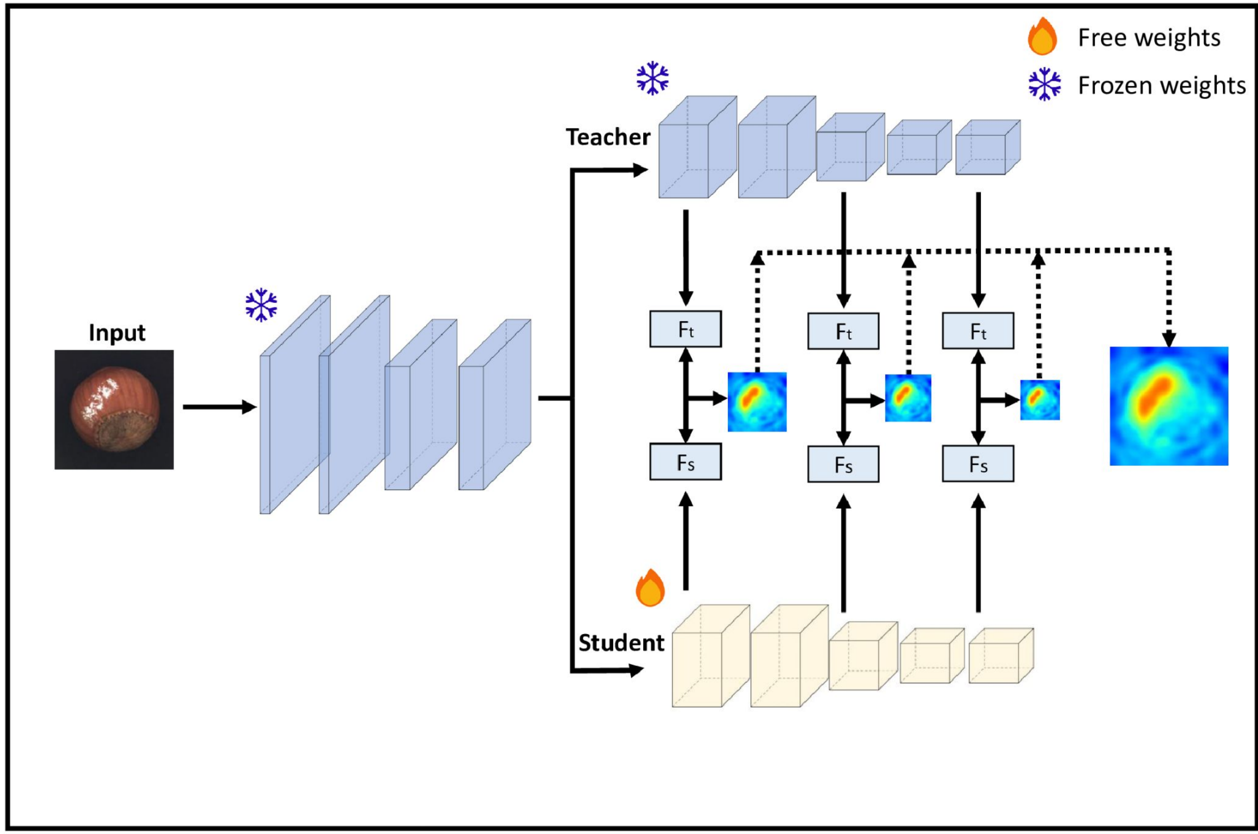
Thesis contribution

PaSTe (Partially Shared Teacher-Student)

STFPM

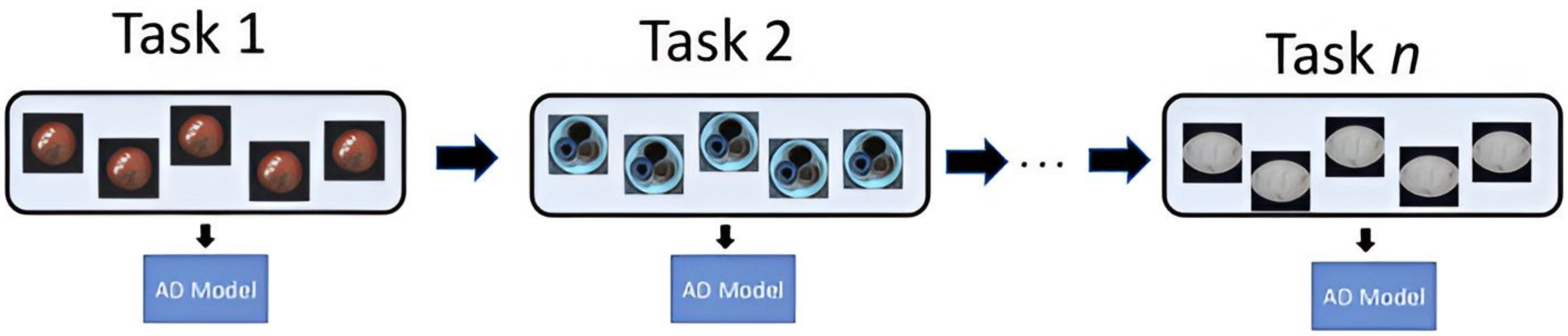


PaSTe

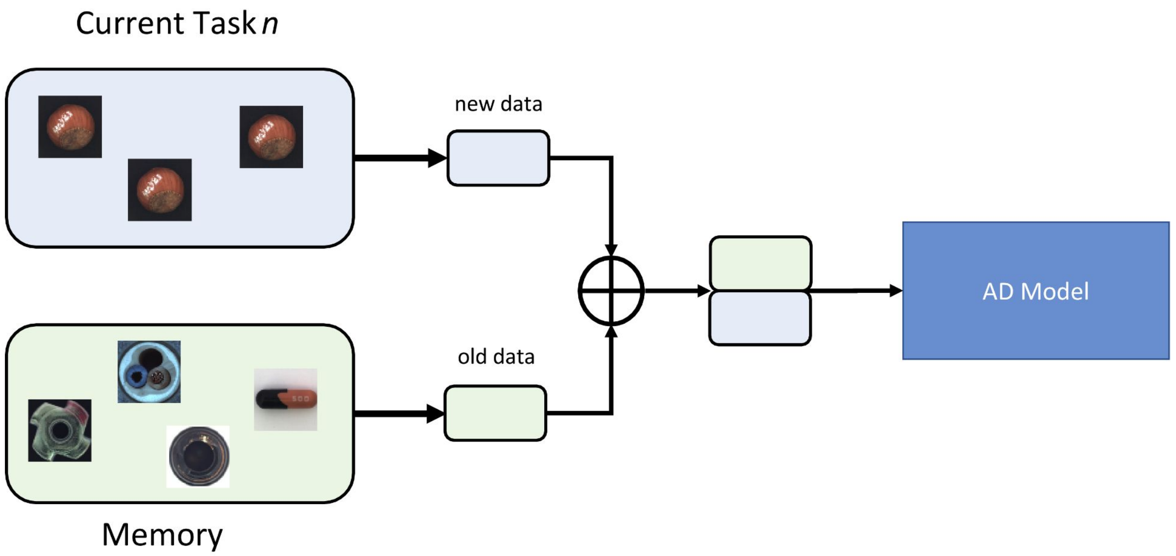


Adapt PaSTe for Continual Learning

CL scenario



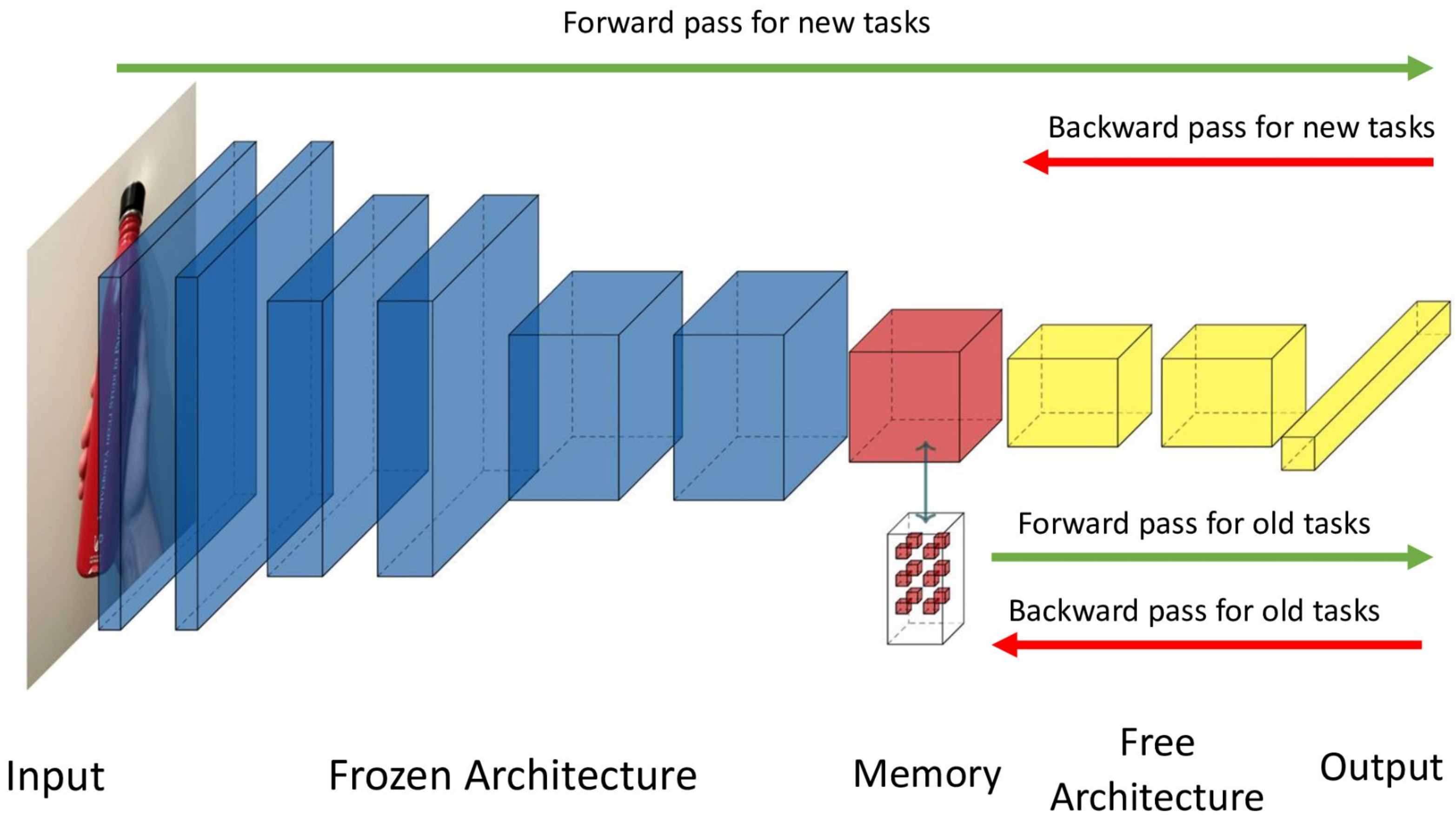
Replay strategy



Compressed Replay

Unavailable for **STFPM**
- **HYP**: First layers are crucial for AD (cannot freeze)

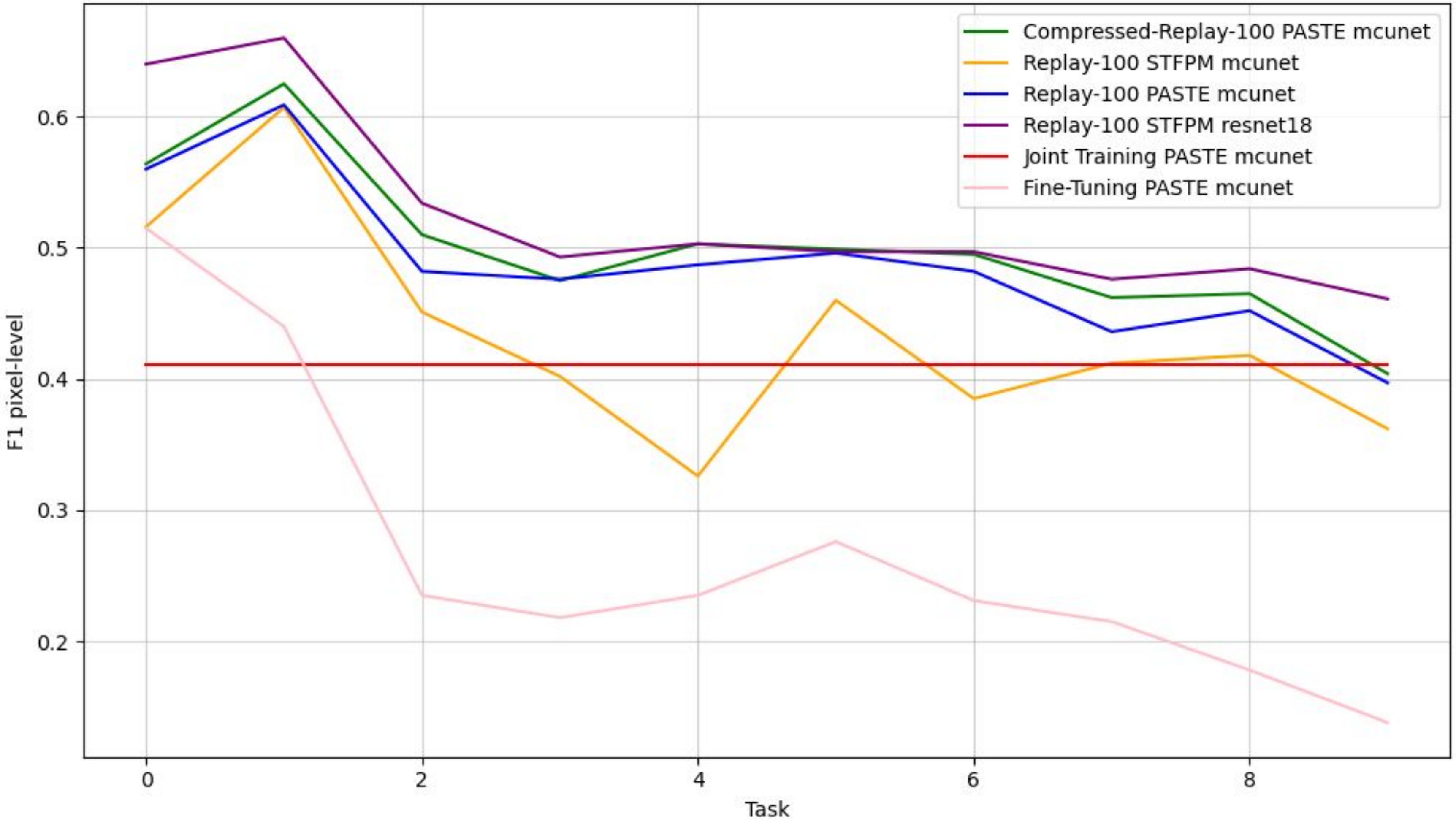
Available for **PaSTe**
- **HYP**: First layers are *not* crucial for AD (shared, frozen)



Continual Learning Strategies Comparison

Università degli Studi di Padova

Each point shows the average performance (**F1 pixel-level**) across all tasks up to that point

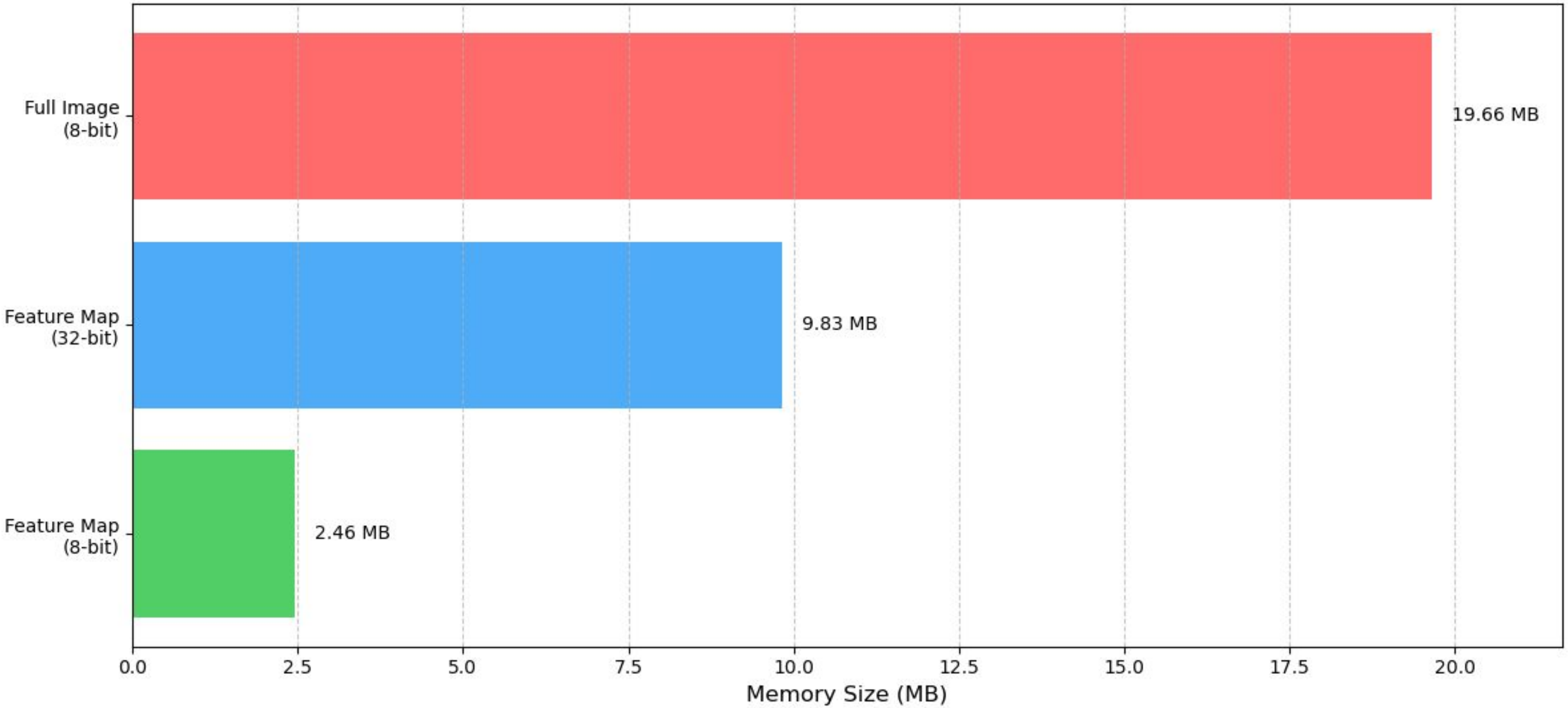


Continual Learning Strategies Comparison

FRAMEWORK		STFPM	STFPM	PaSTe	PaSTe	PaSTe
BACKBONE		ResNet18	MCUNet	MCUNet	MCUNet	MCUNet
CL		Replay	Replay	Finetuning	Replay	Compressed Replay
Image - level	*AUC ROC*	0.87	0.80	0.54	0.80	0.82
	f1	0.90	0.88	0.83	0.87	0.89
Pixel - level	*AUC ROC*	0.92	0.90	0.74	0.90	0.91
	f1	0.46	0.36	0.15	0.40	0.40
	PR AUC	0.42	0.32	0.10	0.36	0.36
	AU PRO	0.84	0.70	0.44	0.74	0.76
Architecture memory [MB]		93.6	1.76	1.68	1.68	1.68
Additional memory [MB]		19.66	19.66	0	19.66	9.83
Average forgetting [%]		0.13	0.20	0.76	0.21	0.19

Feature Quantization

8-bit quantization applied to the feature maps stored in memory (4x memory reduction)



PCA Feature Compression

Università degli Studi di Padova

Performance dropped significantly: critical info required for AD was lost
More advanced feature compression techniques should be explored

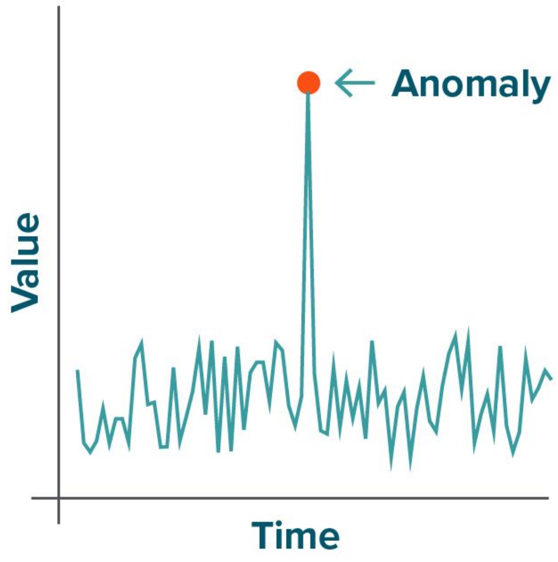
FRAMEWORK		PaSTe		
BACKBONE		MCUNet		
CL		Compressed Replay		
COMPRESSION		None	8-bit Feature Quantization	PCA Feature Compression
Image - level	*AUC ROC*	0.82	0.69	0.67
	f1	0.90	0.86	0.86
Pixel - level	*AUC ROC*	0.91	0.88	0.87
	f1	0.40	0.32	0.31
	PR AUC	0.36	0.26	0.24
	AU PRO	0.76	0.70	0.70
Architecture memory [MB]		1.68	1.68	1.68
Additional memory [MB]		9.83	2.46	$6.89 \rightarrow 8.36^*$
Average forgetting [%]		0.19	0.28	0.29

Thank you!

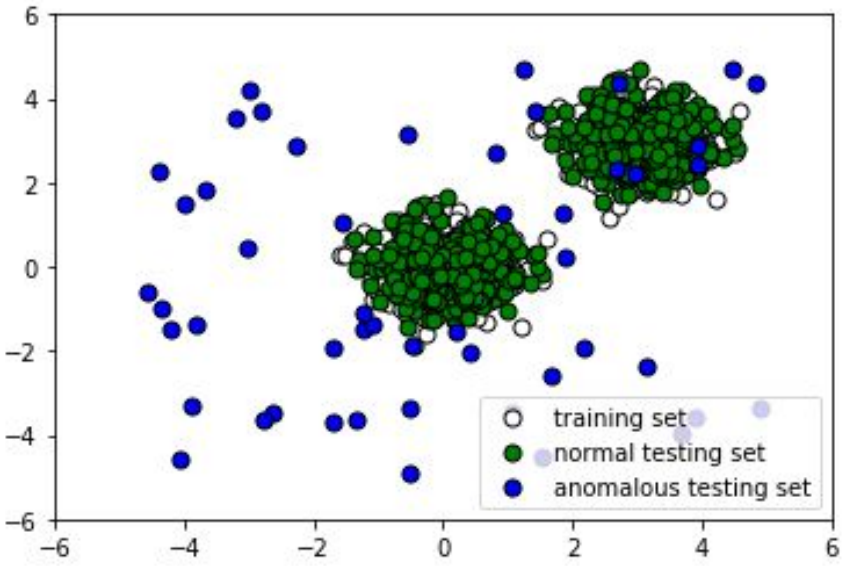
Appendix slides

Anomalies vs Outliers

Aspect	Anomaly	Outlier
Definition	Context-specific deviation	Statistical deviation
Significance	Often meaningful	May be noise or irrelevant
Detection	Requires domain knowledge or models	Can be detected through statistics



Anomalies can be outliers,
not all outliers are anomalies



Visual Anomaly Detection - Strategies

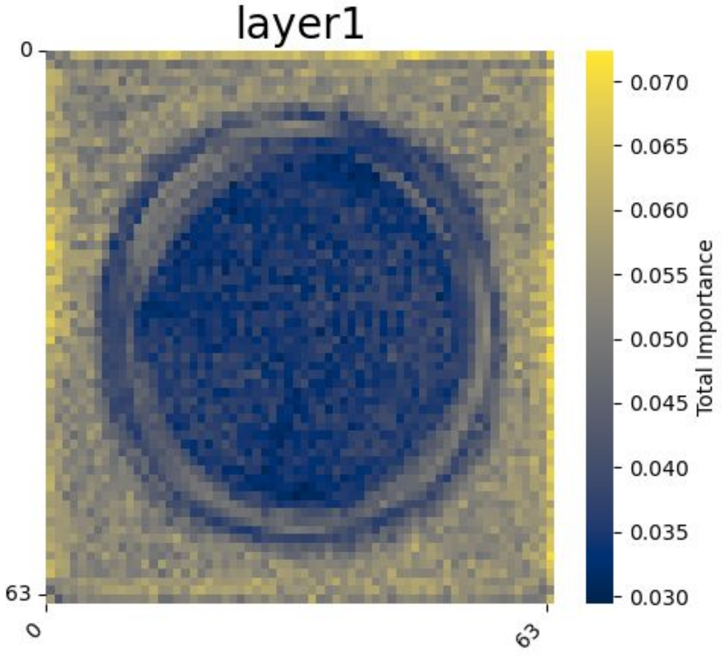
Reconstruction-based

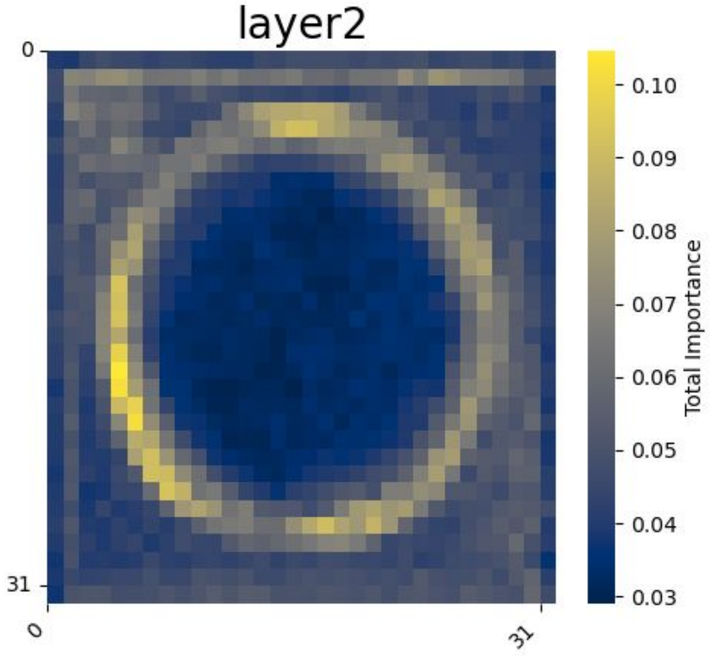
- learn normal image distribution (*training*)
- high reconstruction errors for anomalies (*testing*)
- complex, computationally expensive
- may reconstruct anomalies
- Autoencoders, VAEs, GANs

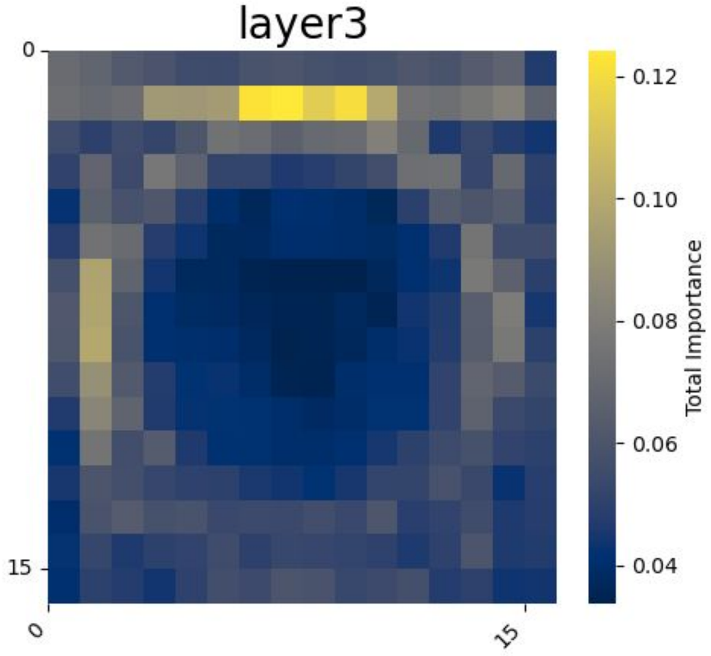
Feature embedding-based

- pre-trained model embeddings
- feature level
- computationally efficient
- better detection accuracy
- Student-Teacher, Memory Bank, Normalizing Flow

Hazelnut (top 1% features)







Pill (top 1% features)