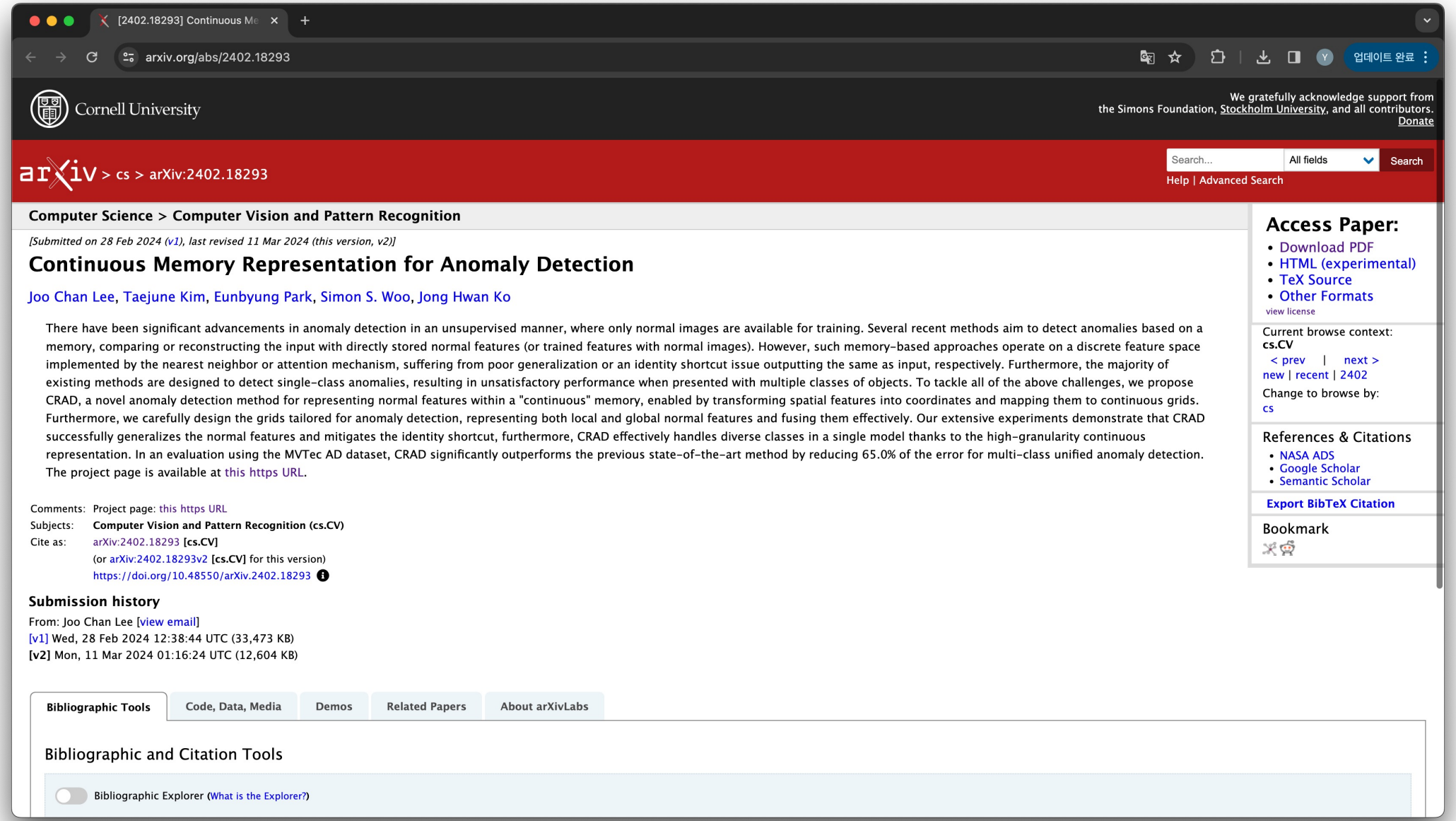
Paper Review

Grid-based Continuous Normal Representation for Anomaly Detection

YeongHyeon Park

Department of Electrical and Computer Engineering

SungKyunKwan University


[2402.18293] Continuous Me
arxiv.org/abs/2402.18293
업데이트 완료
Cornell University
We gratefully acknowledge support from the Simons Foundation, Stockholm University, and all contributors.
Donate
arXiv > cs > arXiv:2402.18293
Search...
All fields
Search
Help | Advanced Search
Computer Science > Computer Vision and Pattern Recognition
[Submitted on 28 Feb 2024 (v1), last revised 11 Mar 2024 (this version, v2)]
Continuous Memory Representation for Anomaly Detection
Joo Chan Lee, Taejune Kim, Eunbyung Park, Simon S. Woo, Jong Hwan Ko
There have been significant advancements in anomaly detection in an unsupervised manner, where only normal images are available for training. Several recent methods aim to detect anomalies based on a memory, comparing or reconstructing the input with directly stored normal features (or trained features with normal images). However, such memory-based approaches operate on a discrete feature space implemented by the nearest neighbor or attention mechanism, suffering from poor generalization or an identity shortcut issue outputting the same as input, respectively. Furthermore, the majority of existing methods are designed to detect single-class anomalies, resulting in unsatisfactory performance when presented with multiple classes of objects. To tackle all of the above challenges, we propose CRAD, a novel anomaly detection method for representing normal features within a "continuous" memory, enabled by transforming spatial features into coordinates and mapping them to continuous grids. Furthermore, we carefully design the grids tailored for anomaly detection, representing both local and global normal features and fusing them effectively. Our extensive experiments demonstrate that CRAD successfully generalizes the normal features and mitigates the identity shortcut, furthermore, CRAD effectively handles diverse classes in a single model thanks to the high-granularity continuous representation. In an evaluation using the MVTec AD dataset, CRAD significantly outperforms the previous state-of-the-art method by reducing 65.0% of the error for multi-class unified anomaly detection. The project page is available at this https URL.
Comments: Project page: this https URL
Subjects: Computer Vision and Pattern Recognition (cs.CV)
Cite as: arXiv:2402.18293 [cs.CV]
(or arXiv:2402.18293v2 [cs.CV] for this version)
https://doi.org/10.48550/arXiv.2402.18293
Submission history
From: Joo Chan Lee [view email]
[v1] Wed, 28 Feb 2024 12:38:44 UTC (33,473 KB)
[v2] Mon, 11 Mar 2024 01:16:24 UTC (12,604 KB)
Access Paper:
Download PDF
HTML (experimental)
TeX Source
Other Formats
view license
Current browse context: cs.CV
< prev | next >
new | recent | 2402
Change to browse by: cs
References & Citations
NASA ADS
Google Scholar
Semantic Scholar
Export BibTeX Citation
Bookmark
Bibliographic Tools
Code, Data, Media
Demos
Related Papers
About arXivLabs
Bibliographic and Citation Tools
Bibliographic Explorer (What is the Explorer?)

Grid-Based Continuous Normal Representation for Anomaly Detection

Joo Chan Lee* Taejune Kim* Eunbyung Park† Simon S. Woo† Jong Hwan Ko†

Sungkyunkwan University, South Korea

* Equal contribution.
† Corresponding authors.

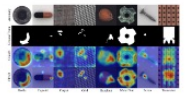
Joo Chan Lee

I am a Ph.D. student in the department of Artificial Intelligence at Sungkyunkwan University, advised by Jong Hwan Ko and Eunbyung Park. My research interest lies in the areas of computer vision, graphics, and machine learning. Currently, I am interested in designing efficient neural fields architecture.

Email / Scholar / LinkedIn / Github

Research

Representative papers are highlighted.

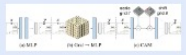


Continuous Memory Representation for Anomaly Detection
Joo Chan Lee*, Taejune Kim*, Eunbyung Park, Simon S. Woo, Jong Hwan Ko
arXiv:2402.18293, 2024
[Project page] [Paper] [Code]

A novel approach to learning normal representation in continuous feature space for anomaly detection.

Compact 3D Gaussian Representation for Radiance Field
Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, Eunbyung Park
CVPR, 2024
[Project page] [Paper] [Code]

A comprehensive framework for 3D scene representation, achieving high performance, fast training, compactness, and real-time rendering.

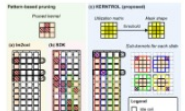


Coordinate-Aware Modulation for Neural Fields
Joo Chan Lee, Daniel Rho, Seungtae Nam, Jong Hwan Ko, Eunbyung Park
ICLR, 2024 **(Spotlight)**
[Project page] [Paper] [Code]

Injecting spectral bias-free grid representations into the intermediate features of the MLP achieves high performance with compactness.

FFNeRV: Flow-Guided Frame-Wise Neural Representations for Videos
Joo Chan Lee, Daniel Rho, Jong Hwan Ko, Eunbyung Park
ACM MM, 2023
[Project page] [Paper] [Code]

Incorporating flow information into frame-wise representations to exploit the temporal redundancy across the frames in videos.



KERNTROL: Kernel Shape Control Toward Ultimate Memory Utilization for In-Memory Convolutional Weight Mapping
IEEE TCAS-I, 2024 [Paper]
Kernel Shape Control for Row-Efficient Convolution on Processing-In-Memory Arrays
ICCAD, 2023 [Paper] [Code]
Johnny Rhe, Kang Eun Jeon, **Joo Chan Lee**, Seongmoon Jeong, Jong Hwan Ko

https://maincold2.github.io/
https://tae-mo.github.io/

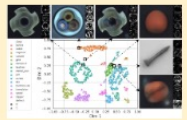
Taejune Kim

Robotics Lab, Hyundai Motor Company
I completed BSc and MSc in Computer Science and Engineering

Email / LinkedIn / Github

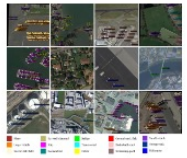
Research

I'm interested in computer vision, anomaly detection, and object detection. Representative papers are highlighted.



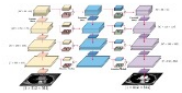
Continuous Memory Representation for Anomaly Detection
Joo Chan Lee*, **Taejune Kim***, Eunbyung Park, Simon S. Woo, Jong Hwan Ko
under review
project page / code

Unified framework for unsupervised anomaly detection.



Rotated-DETR: an End-to-End Transformer-based Oriented Object Detector for Aerial Images
Jinbeom Kim*, Giljun Lee*, **Taejune Kim**, Simon S. Woo
SAC, 2023

Inferencing oriented bounding box using Deformable-DETR framework.

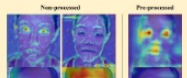


MGCMA: Multi-scale Generator with Channel-wise Mask Attention to generate Synthetic Contrast-enhanced Chest Computed Tomography
Jeongho Kim, Yun-Gyoo Lee, Donggeun Ko, **Taejune Kim**, Soo-Youn Ham, Simon S. Woo
SAC, 2023

Synthesizing contrast-enhanced style on non-contrast CT scans.

A^2: Adaptive Augmentation for Effectively Mitigating Dataset Bias
Jaeju An, **Taejune Kim**, Donggeun Ko, Sangyup Lee, Simon S. Woo
ACCV, 2022

Mitigating dataset bias through domain adaptation.



Evading Deepfake Detectors via High Quality Face Pre-Processing Methods
Jeongho Kim*, **Taejune Kim***, Jeonghyeon Kim, Simon S. Woo
ICPR, 2022

AD Approaches

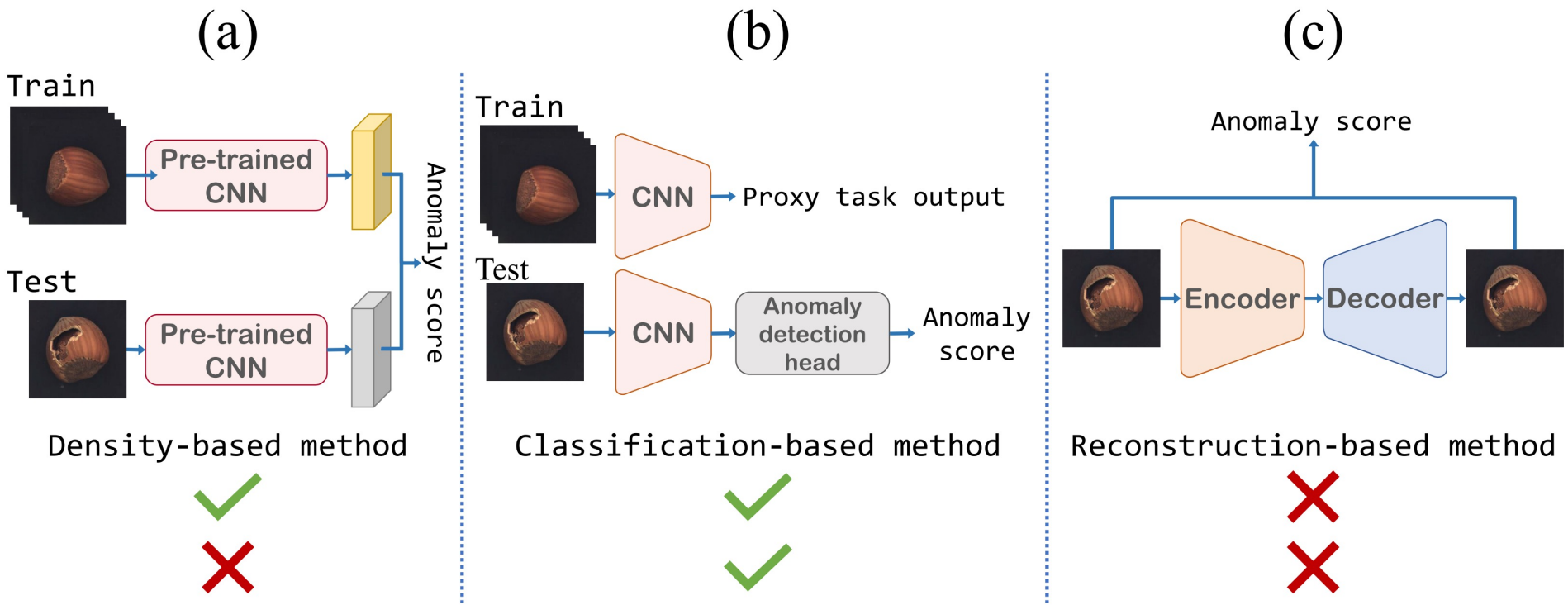


Fig. 2. Pipeline illustrations of three kinds of unsupervised anomaly detection methods in column. Bottom two rows indicate whether *Pretrained Model* and *Extra Training Data* are used for each kind of method.

AD Approaches

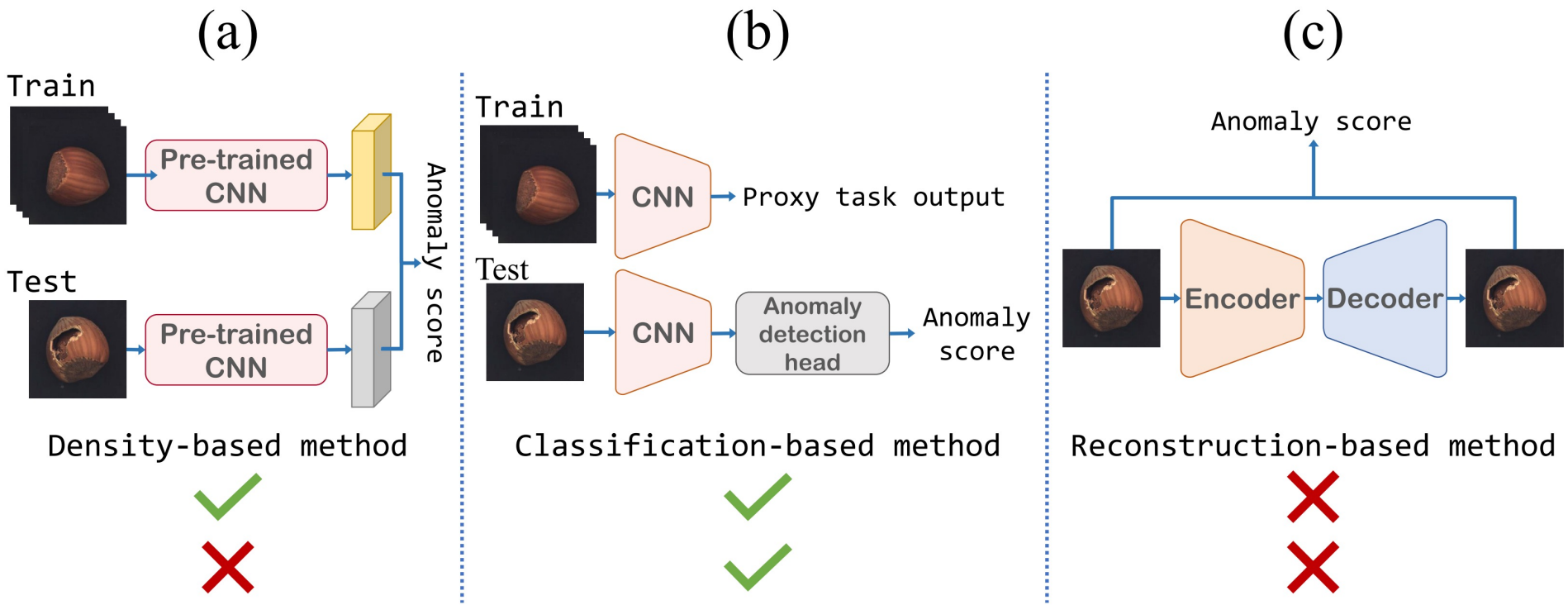


Fig. 2. Pipeline illustrations of three kinds of unsupervised anomaly detection methods in column. Bottom two rows indicate whether *Pretrained Model* and *Extra Training Data* are used for each kind of method.

Feature-level detection

Image-level detection

GRAD

Grid Representation for Anomaly Detection

Need for limited generalization ability

Another line of work [13, 15, 25] focuses on producing generalized normal features. Unlike the aforementioned approaches that use the nearest neighbor technique, these methods combine multiple normal features from the memory using an attention mechanism (i.e., referring to multiple discrete features), given a normal or abnormal input (Fig. 1(b)). They assume that the model always generates normal features, regardless of whether the inputs are normal or abnormal, thus anomalous regions can be detected based on the disparity between the inputs and outputs. Because these models gather diverse normal features from the memory via attention, they have exhibited increased robustness to test data, leading to improved generalization performance. However, such strength may turn into a drawback when testing abnormal inputs. If these inputs can be reconstructed using a combination of normal features, the model might ultimately generate an output that is identical to the abnormal input. This issue, termed as an identity shortcut (IS) by UniAD [40], prevents the models from detecting anomalies due to the lack of disparity between the abnormal input and produced output.

Concept

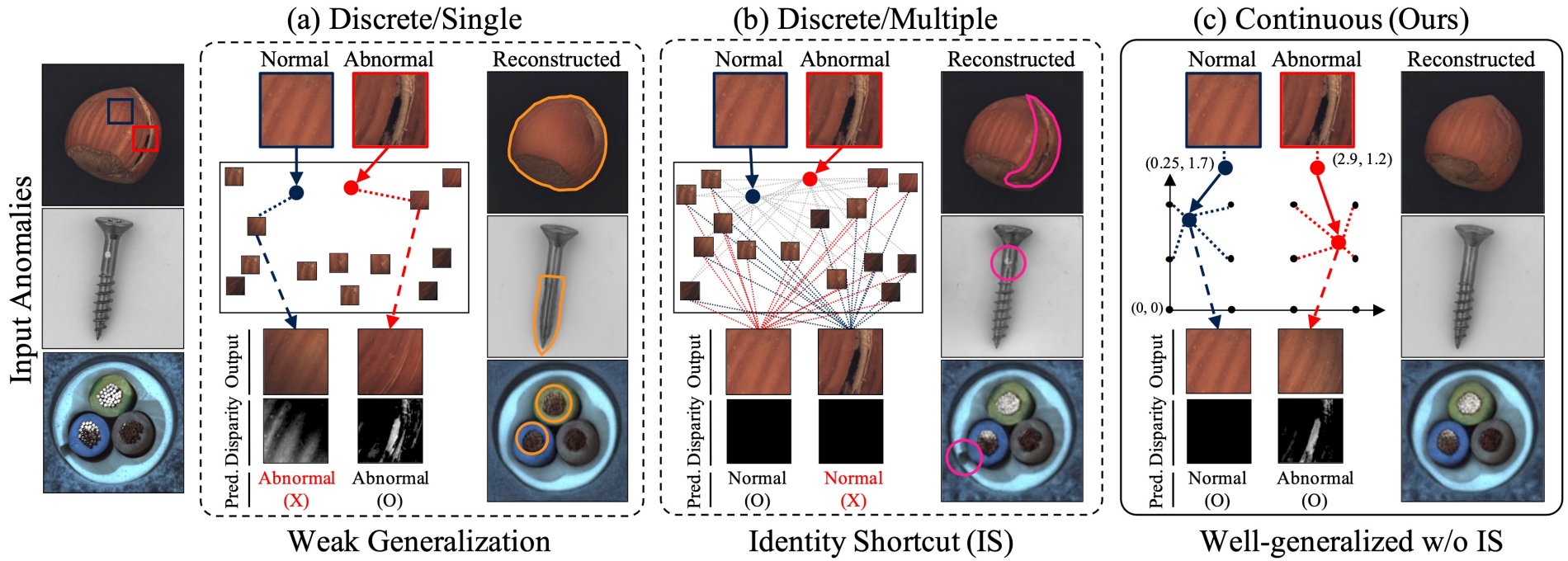


Fig. 1: Conceptual diagram and qualitative results of existing methods and ours. (a) and (b) use single and multiple normal features in a discrete space, respectively, while our method (c) exploits continuous feature space. We visualize the anomaly detection process with the normal (navy) and abnormal (red) patches of the top-left reference image. 'Pred.' indicates the prediction based on the disparity, and wrong predictions are marked as (X) with red color. We present the reconstruction results based on the reference abnormal images.

Summaries

Motivation and solution

- **Motivation**: Integration of anomaly detection model for multiple tasks
  - Train a model for a total of all tasks : multi-class AD with a single model

- **Solution**: Transform all input features into specific coordinated of continuous space
  - Force to move input feature into a center of the feature space

Contributions (Authors said)

- Representing normal features within a continuous feature space by grid operation

- Design the grid for representing both local and global normal features

- Successfully generalizes normal features and mitigates identity shortcut

Remaining Questions

- Why does the GRAD utilize third and fourth stage feature maps?

- Why authors show image reconstruction results?
  - Where is the explanation of the process of obtaining image reconstruction results?

- GRAD seem to be detecting anomalies by blurring the high-frequency components
  - that appear in the defective samples.

Summaries

Motivation and solution

- **Motivation**: Integration of anomaly detection model for multiple tasks
  - Train a model for a total of all tasks : multi-class AD with a single model
- **Solution**: Transform all input features into specific coordinated of continuous space
  - Force to move input feature into a center of the feature space

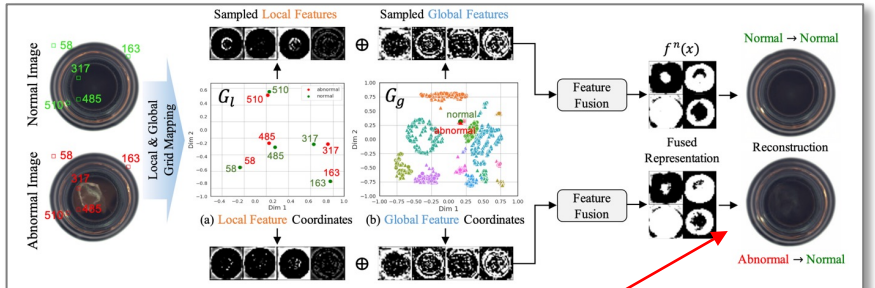


Fig. 3: Visualization of GRAD's pipeline. Each marker in (a) corresponds to the patch on the left image that has the same number and color. Each marker in (b) corresponds to a single image from the test dataset, where different colors represent distinct classes, and circles and triangles denote the normal and abnormal images, respectively. 'Dim 1' and 'Dim 2' are the two dimensions of 2D grids.

Contributions (Authors said)

- Representing normal features within a continuous feature space by grid operation
- Design the grid for representing both local and global normal features
- Successfully generalizes normal features and mitigates identity shortcut

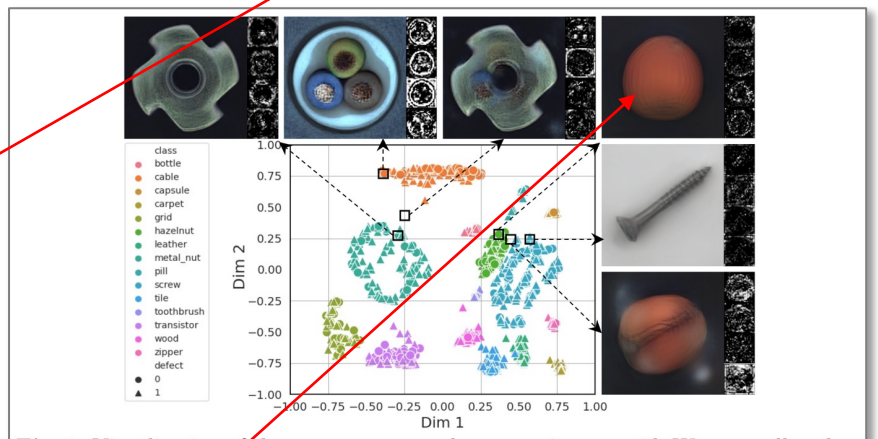


Fig. 4: Visualization of the contents mapped at a continuous grid. We manually select six global coordinates and visualize the corresponding sampled normal features.

Remaining Questions

- Why does the GRAD utilize third and fourth stage feature maps?
- Why authors show image reconstruction results?
  - Where is the explanation of the process of obtaining image reconstruction results?
- GRAD seem to be detecting anomalies by blurring the high-frequency components
  - that appear in the defective samples.

Anyway, the AD performance is good.

Overview

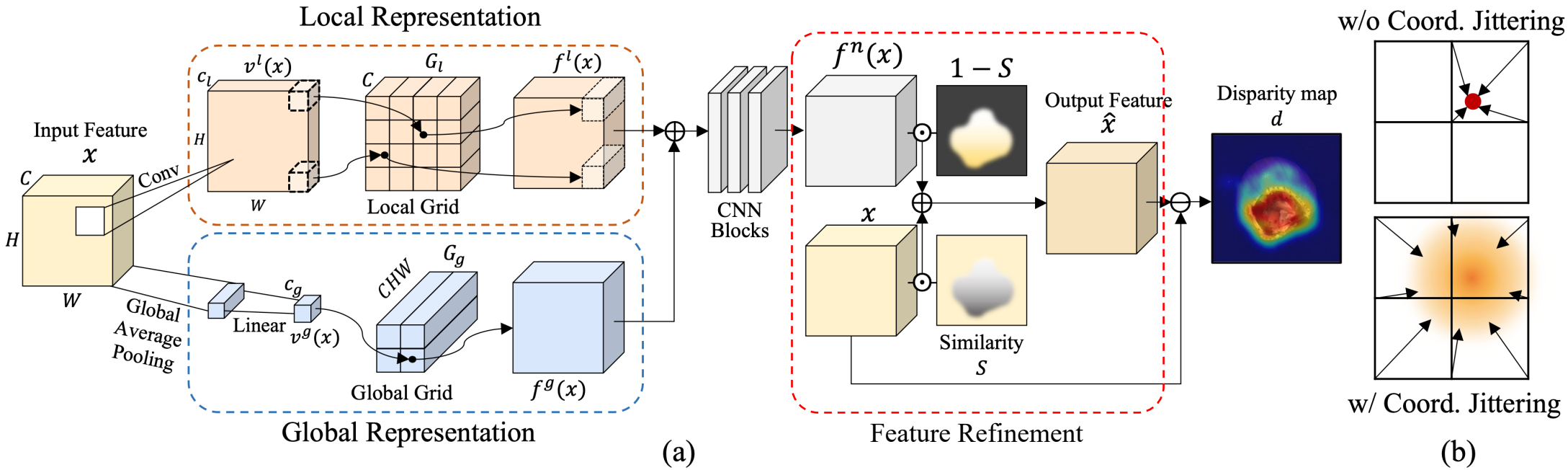
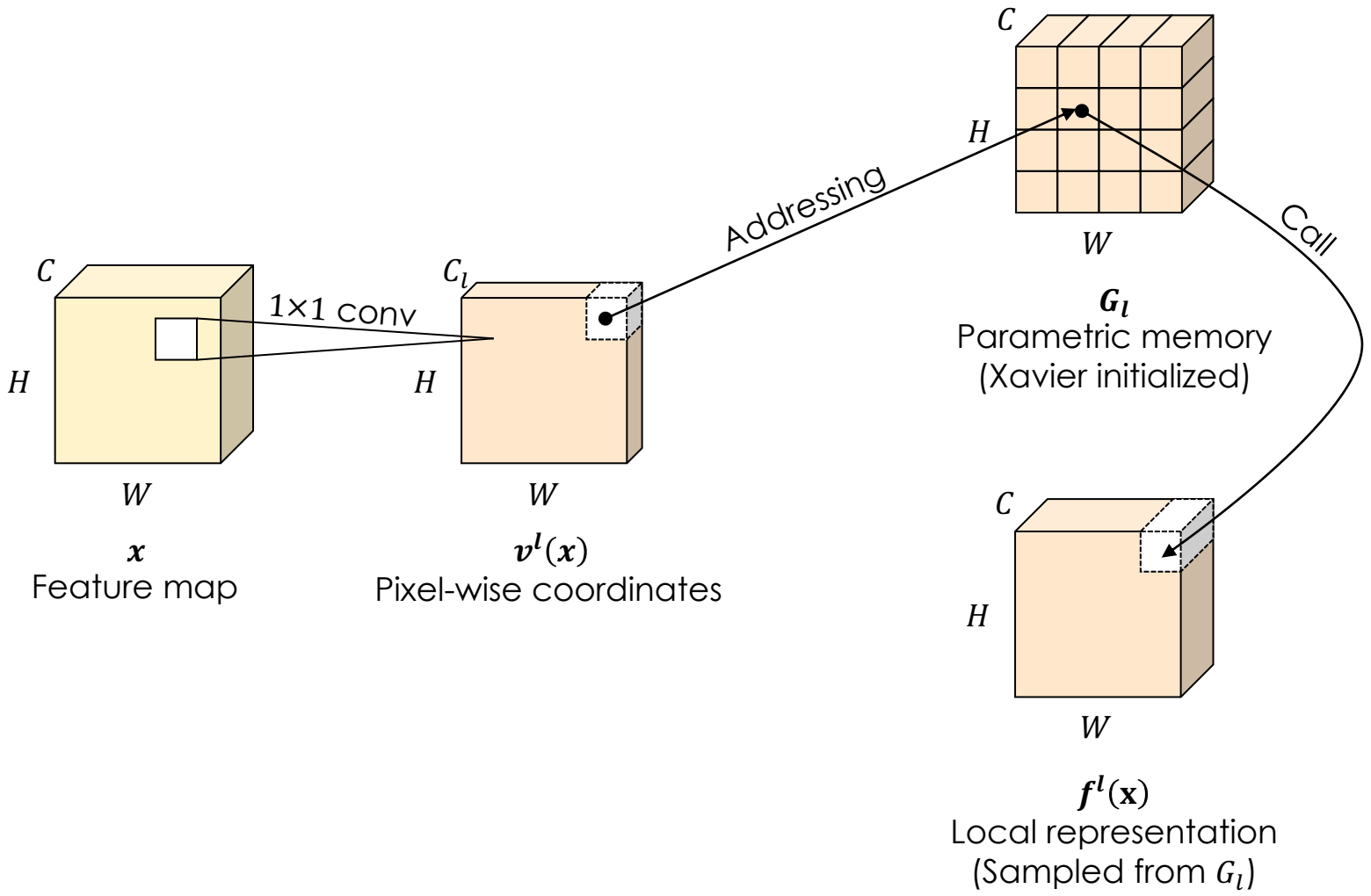


Fig. 2: (a) The detailed architecture of GRAD and (b) visualization of coordinate jittering. The input x is firstly transformed into pixel-wise and feature-wise coordinates. After the normal features are sampled from local and global representations, they are fused by CNN blocks. The final reconstruction is acquired through the proposed feature refinement process.

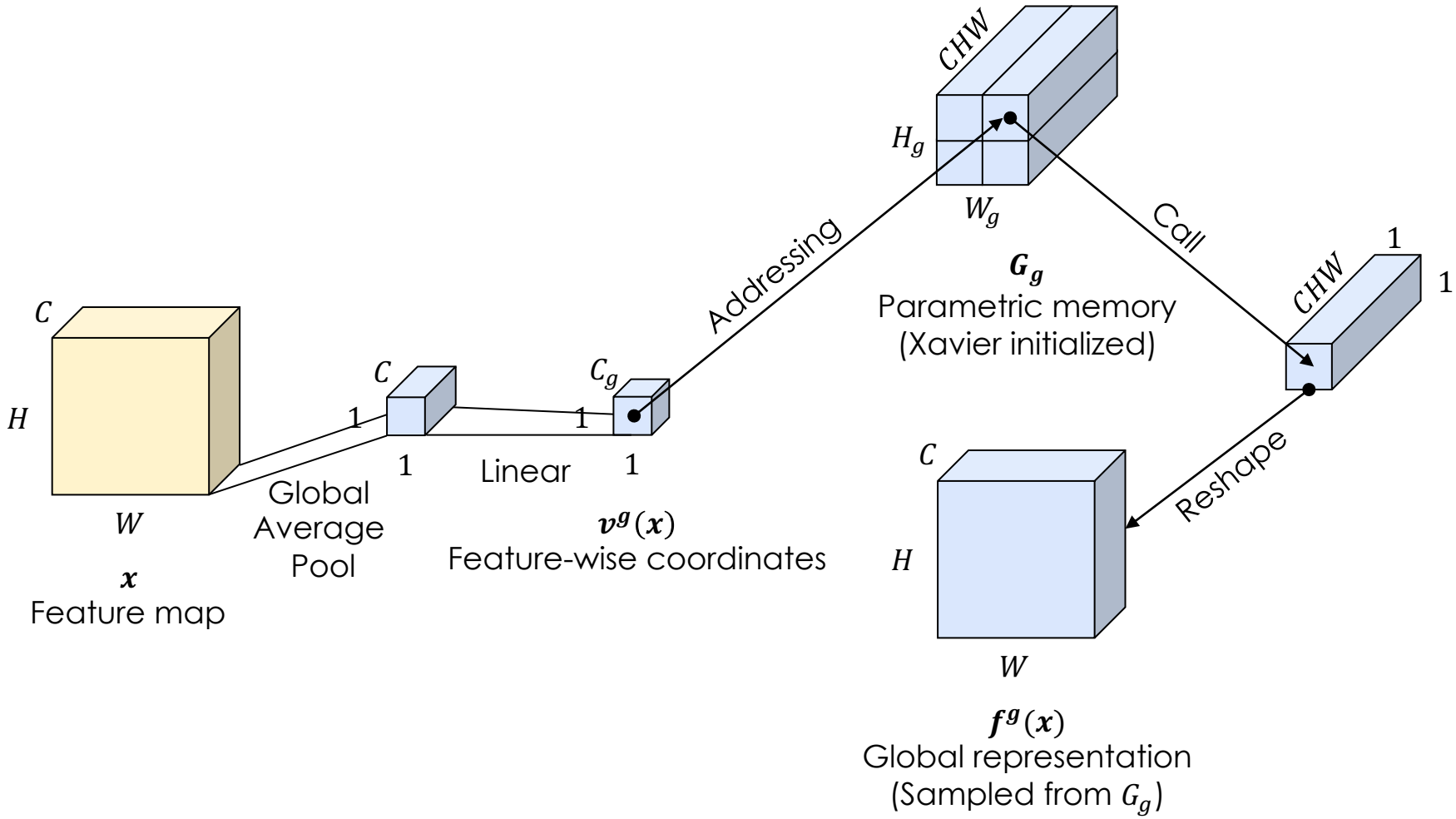
Local Representation

* $C_l = 2$



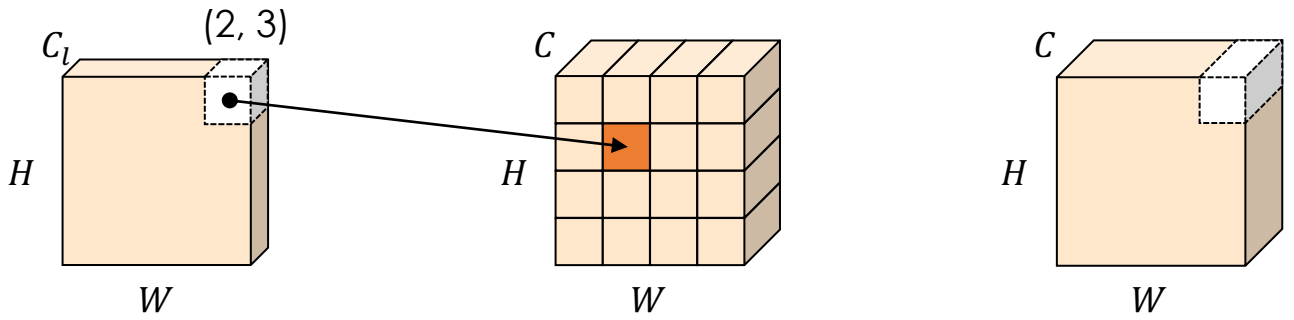
Global Representation

* $C_g = 2$

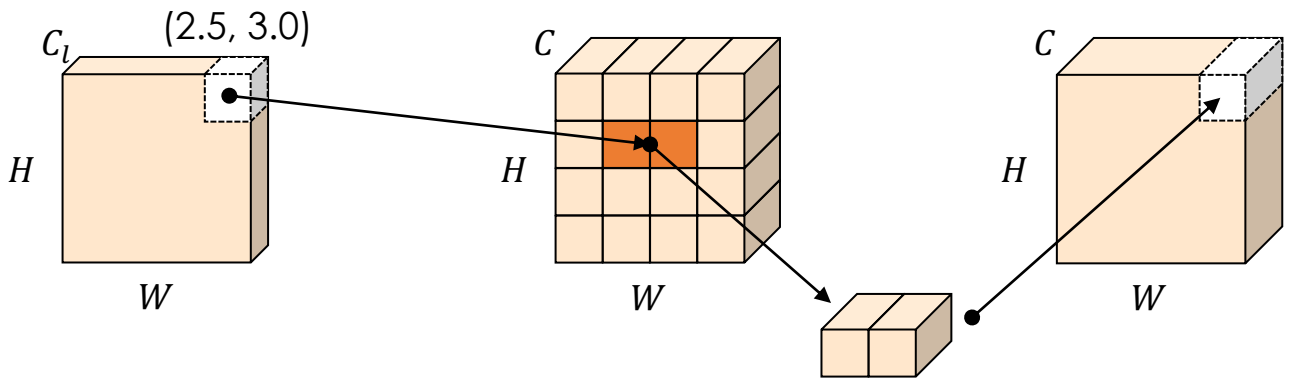


Grid Operation

Integer Coordinates (Discrete)



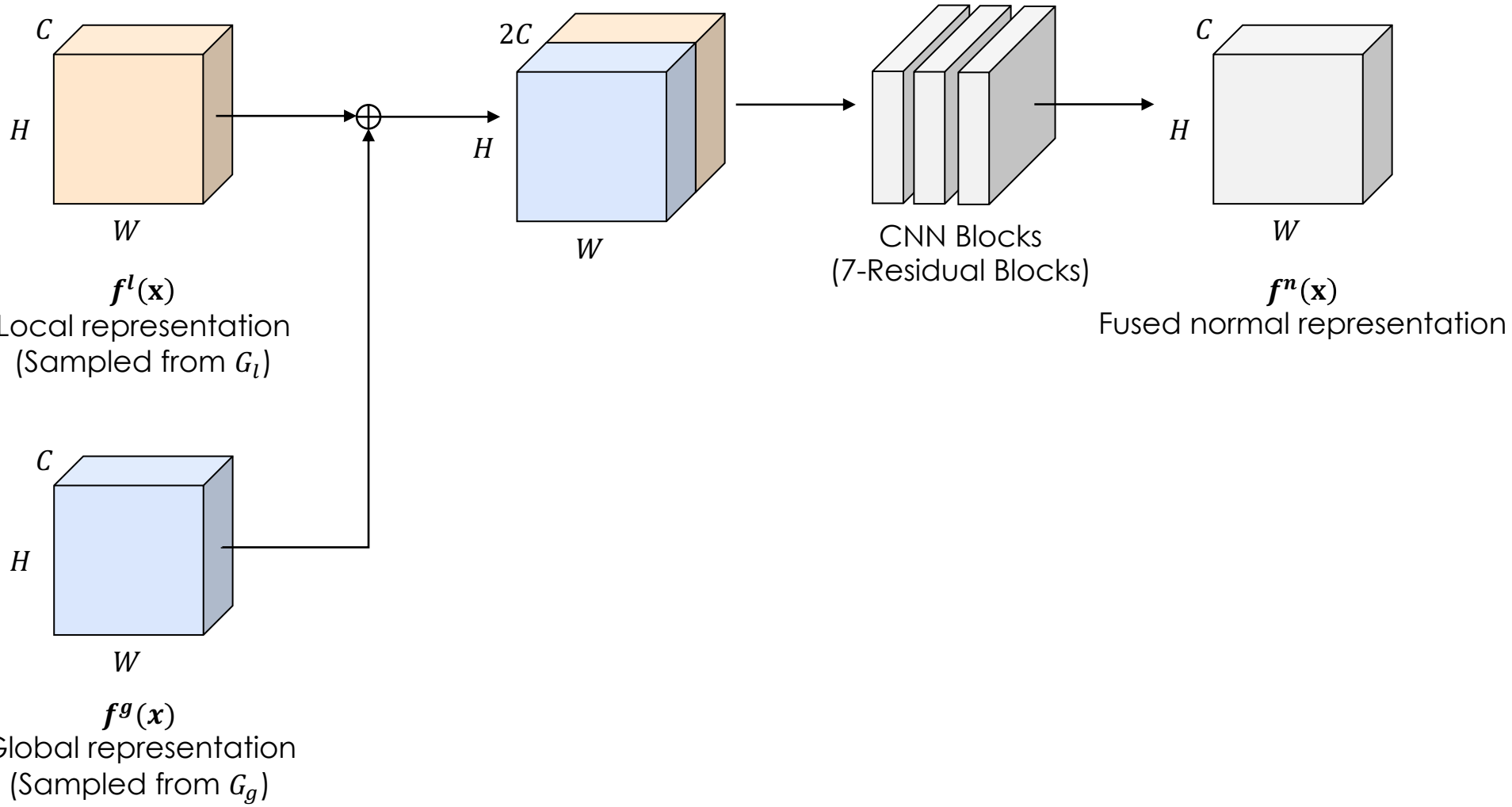
Floating-point number Coordinates (Continuous)



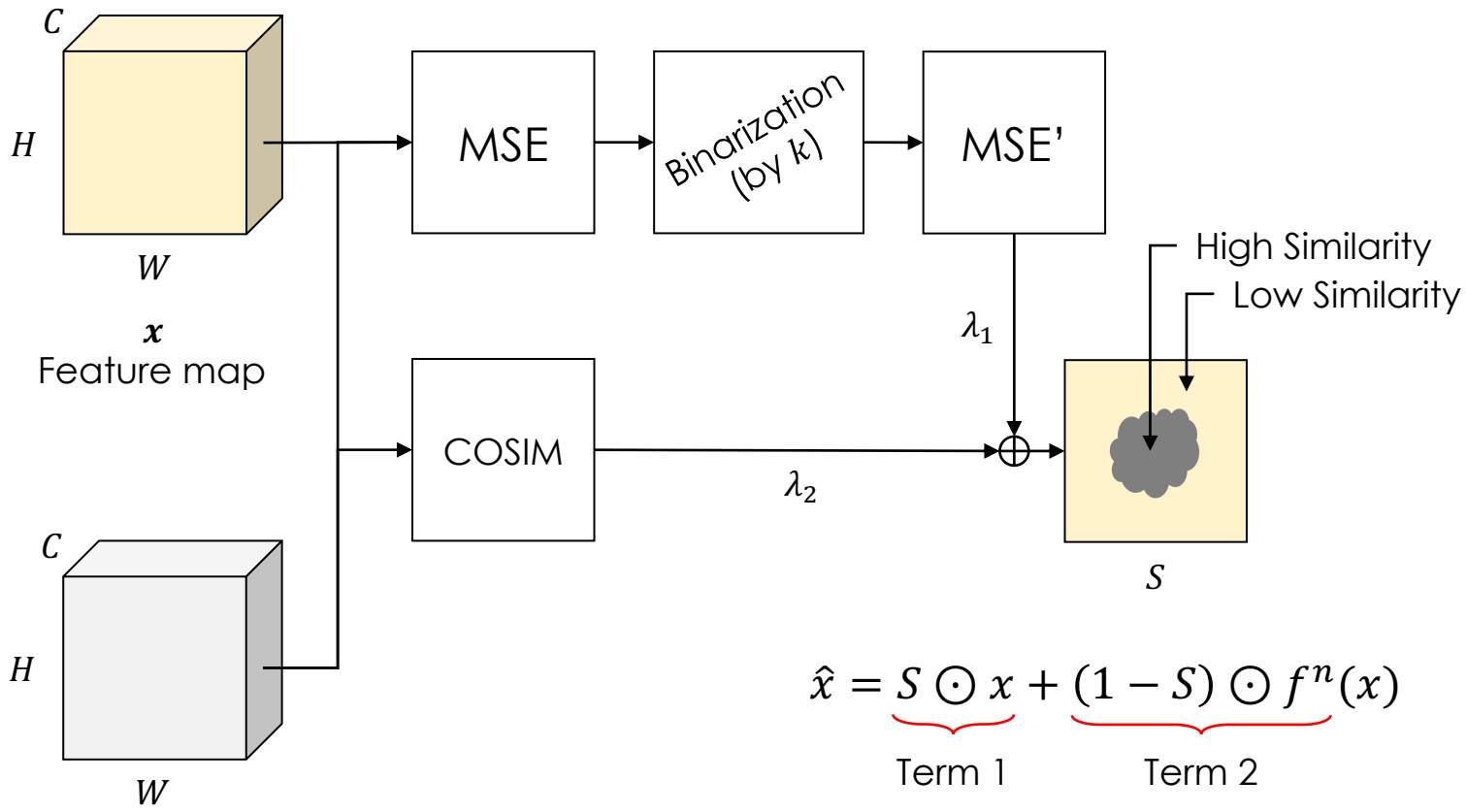
$$|v - n|G[m] + |v - m|G[n]$$

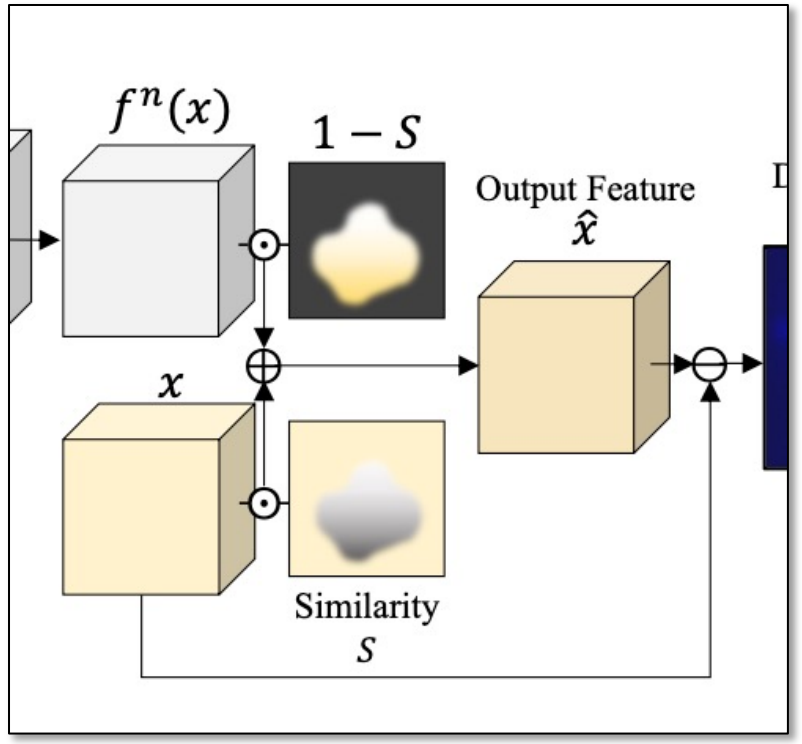
$$\Rightarrow |2.5 - 3|G[2] + |2.5 - 2|G[3]$$

Fused representation



Feature refinement



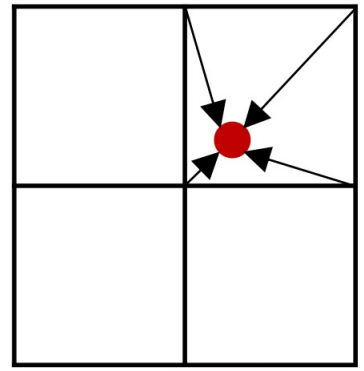


$$\hat{x} = \underbrace{S \odot x}_{\text{Term 1}} + \underbrace{(1 - S) \odot f^n(x)}_{\text{Term 2}}$$

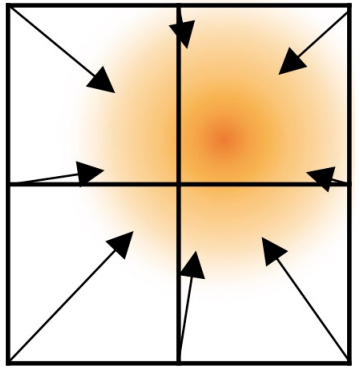
- **Term 1**: Pass normal feature directly
- **Term 2**: Block abnormal feature by transform normal feature
  - Same as replacing abnormal features into normal form

* Regions of S with low values represent regions of x that includes abnormal features

Training and Inference



(w/o coord. jit.)



(w coord. jit.)

Training

- Update entire model in an *end-to-end manner*
- Initialize grids by *Xavier initializer*
- Update parameters by *MSE loss* between x and $\hat{x}$
- Utilize *coordinate jittering* to achieve a more generalized grid representation

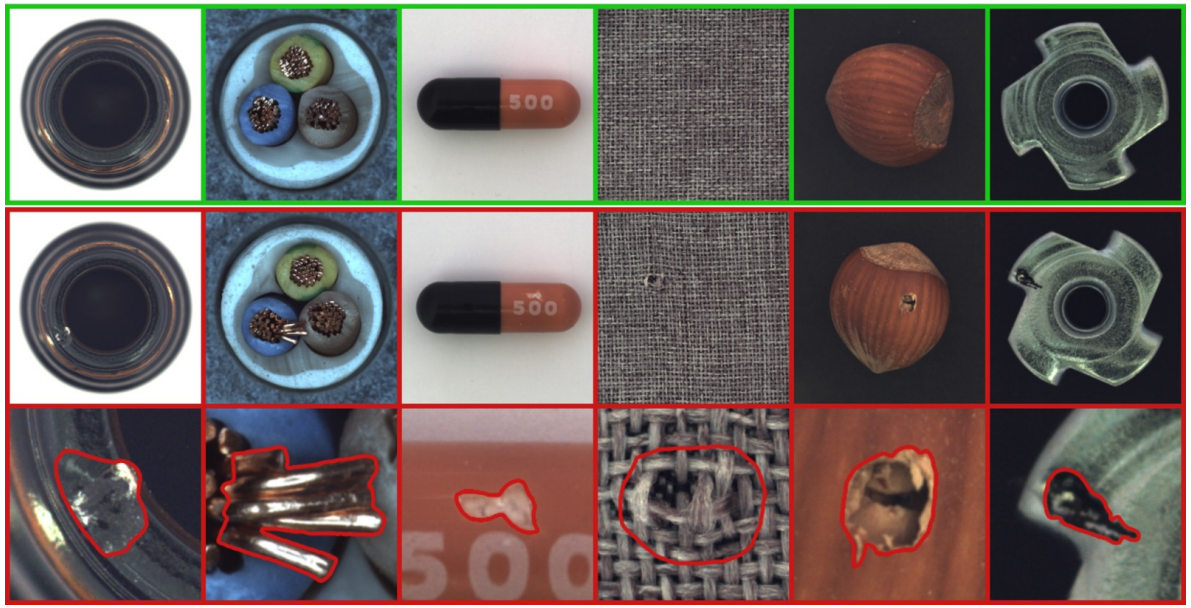
Inference

- Measure anomaly score by pixel-wise *L2 Norm*
- Use maximum value of average pooled pixel-wise *L2 Norm as a image-level anomaly score*

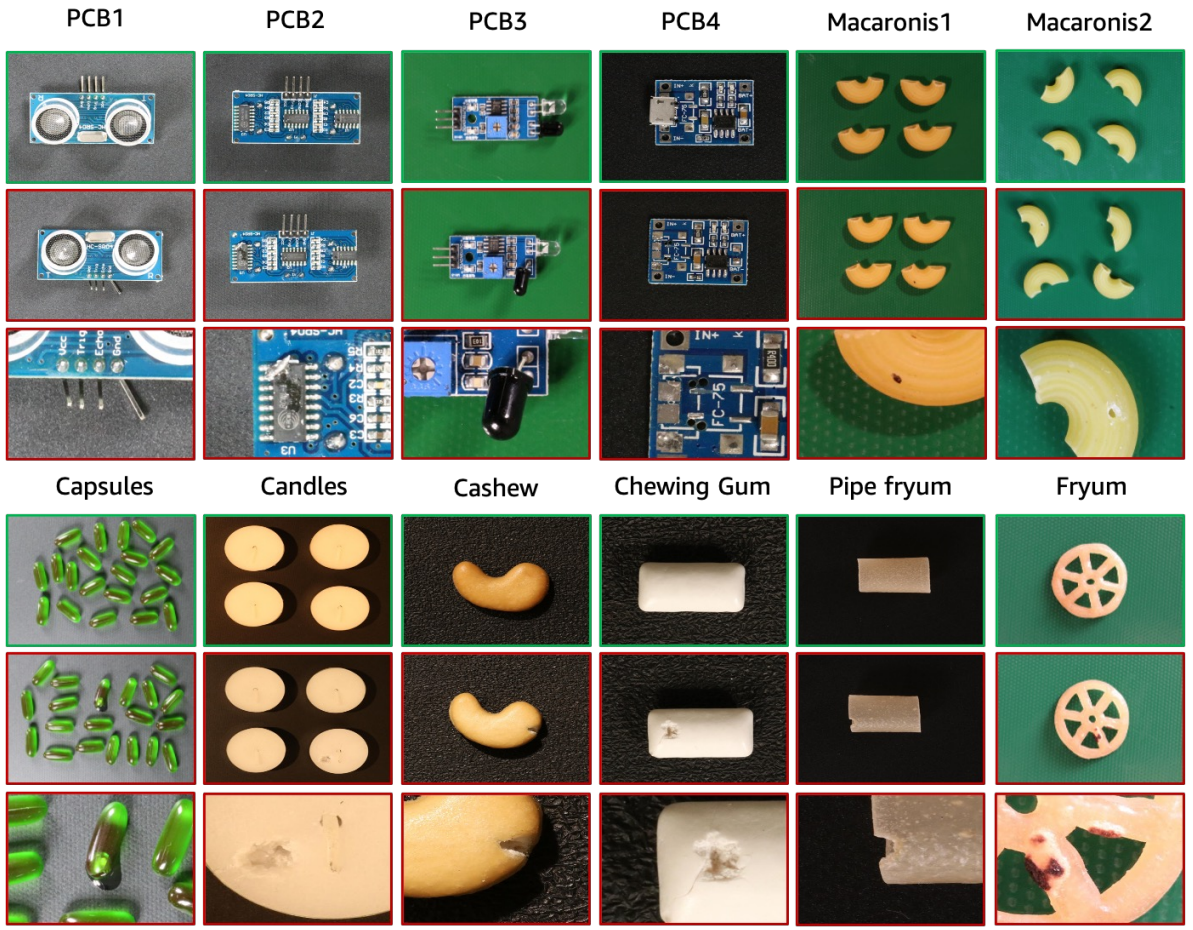
Experiments

Paul Bergmann, et al. "MVTec AD-A comprehensive real-world dataset for unsupervised anomaly detection." CVPR. 2019.
Yang Zou, et al. "Spot-the-difference self-supervised pre-training for anomaly detection and segmentation." ECCV. 2022.

Datasets



MVTec AD



VisA

Pipeline of GRAD

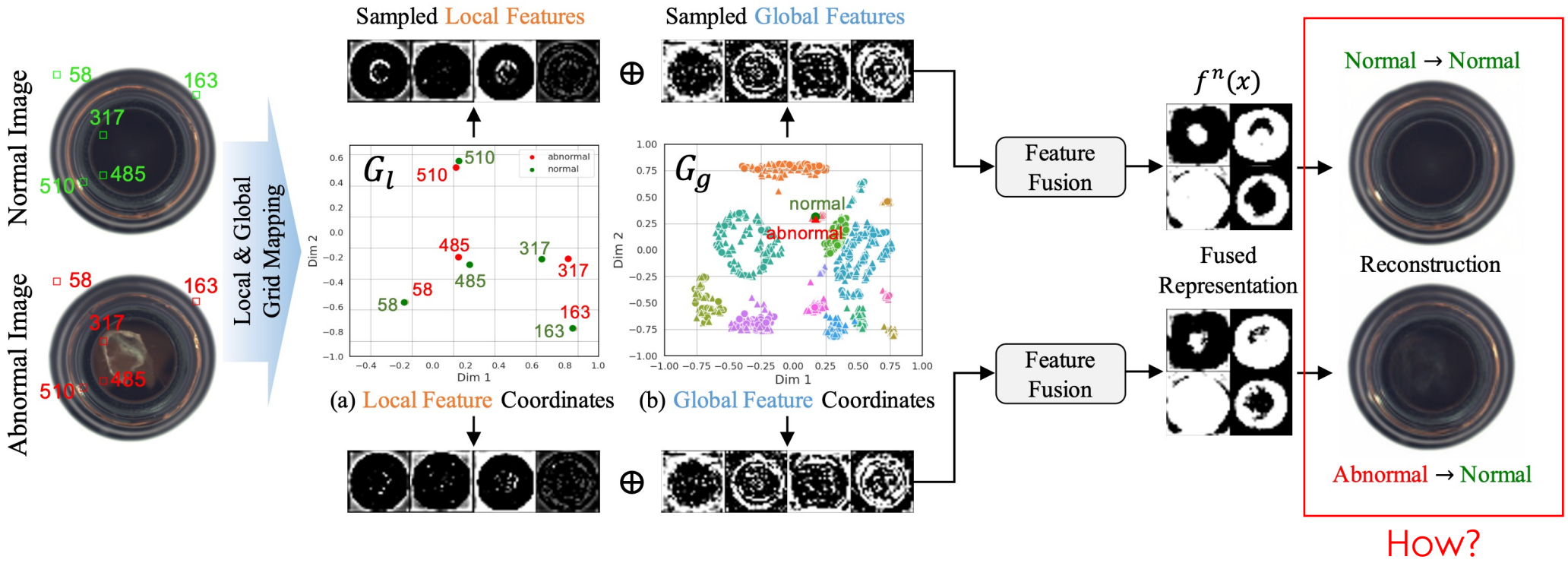


Fig. 3: Visualization of GRAD's pipeline. Each marker in (a) corresponds to the patch on the left image that has the same number and color. Each marker in (b) corresponds to a single image from the test dataset, where different colors represent distinct classes, and circles and triangles denote the normal and abnormal images, respectively. 'Dim 1' and 'Dim 2' are the two dimensions of 2D grids.

Pipeline of GRAD

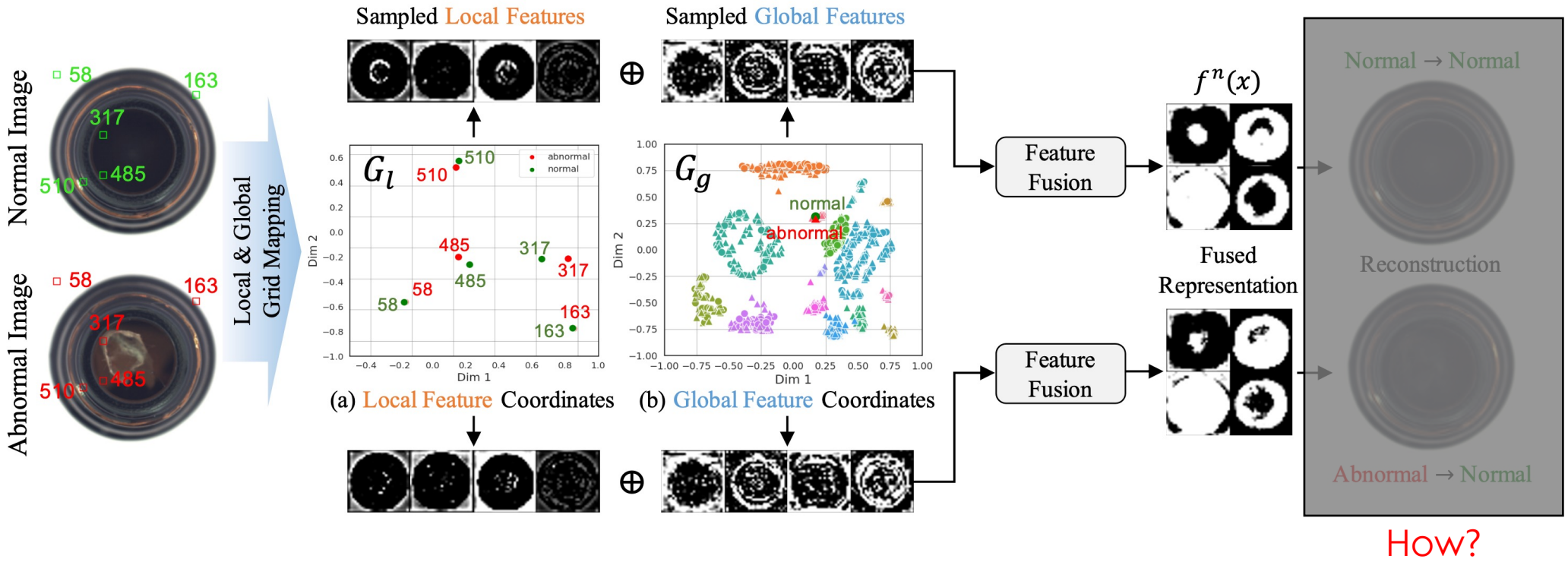


Fig. 3: Visualization of GRAD's pipeline. Each marker in (a) corresponds to the patch on the left image that has the same number and color. Each marker in (b) corresponds to a single image from the test dataset, where different colors represent distinct classes, and circles and triangles denote the normal and abnormal images, respectively. 'Dim 1' and 'Dim 2' are the two dimensions of 2D grids.

Latent representation of GRAD

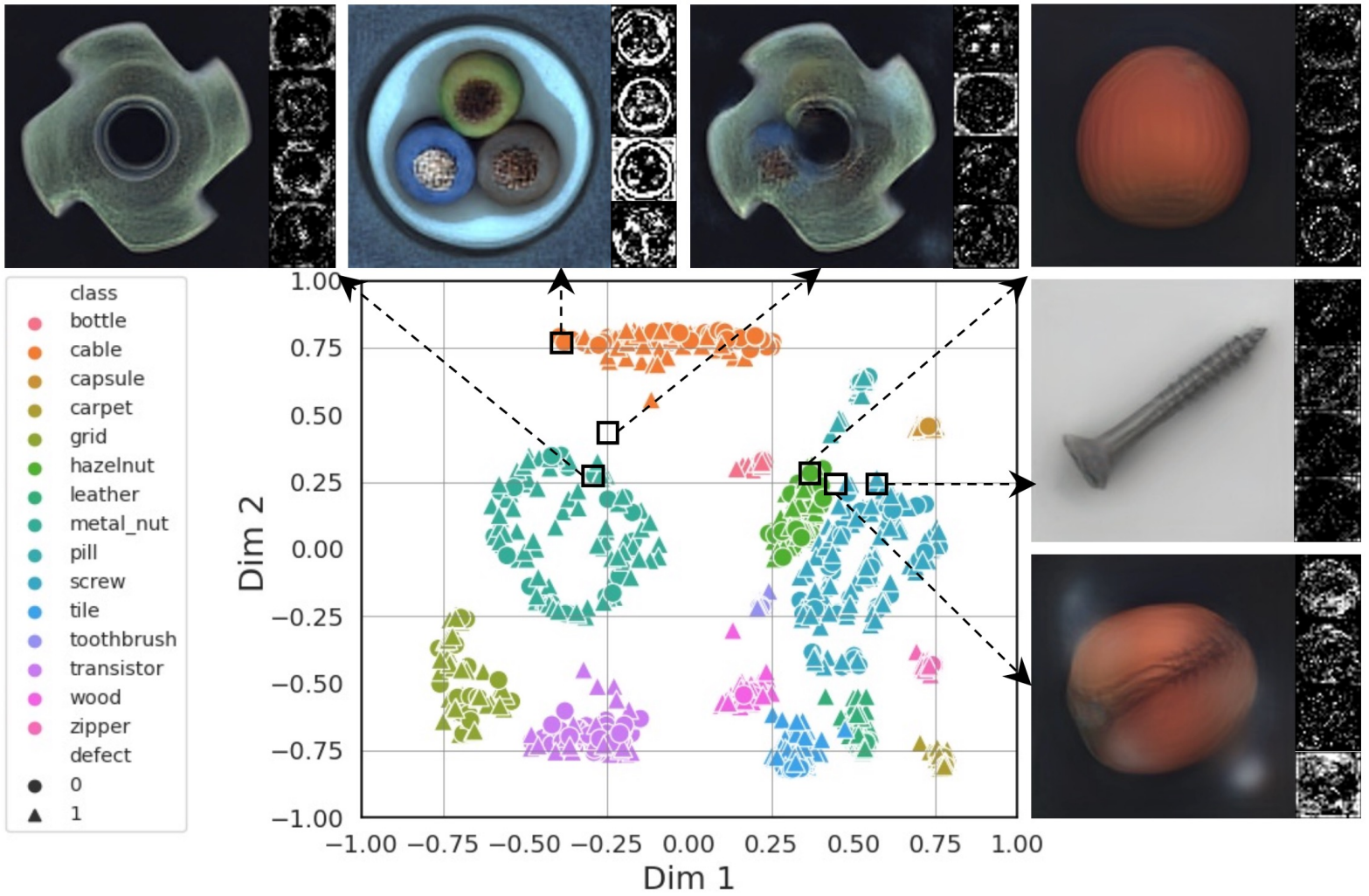
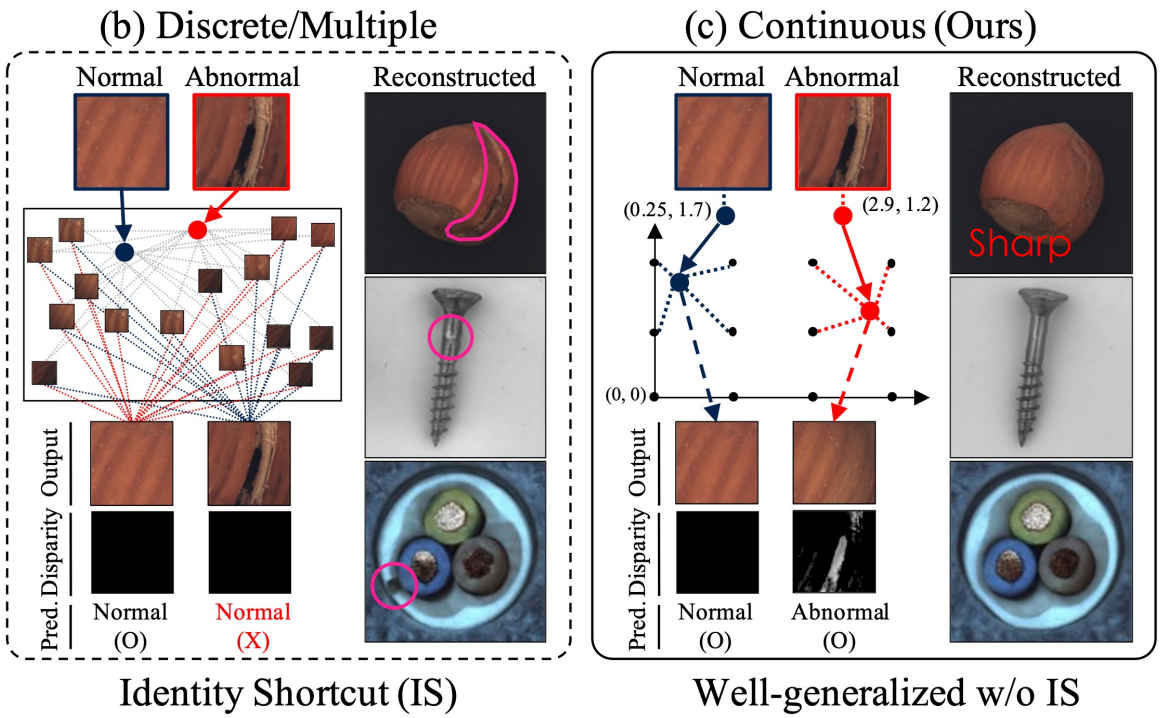
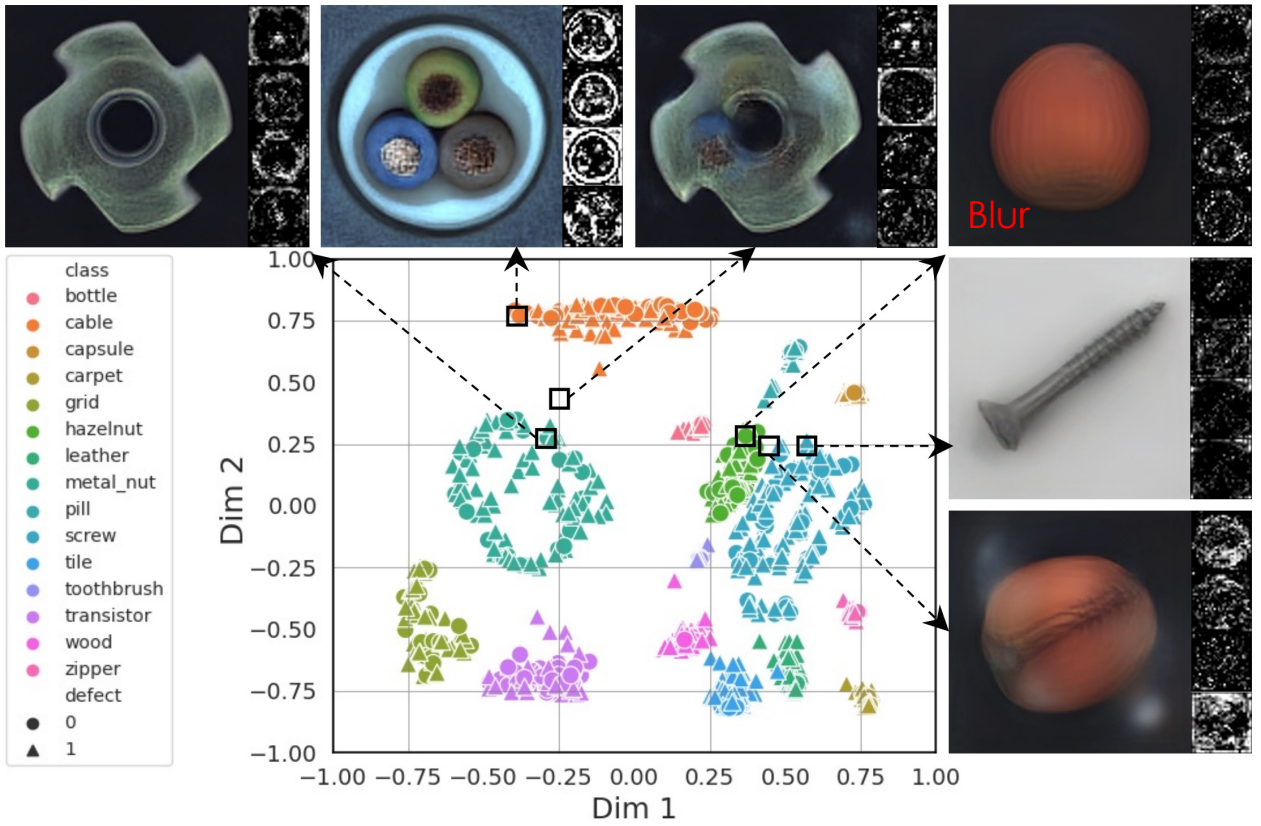


Fig. 4: Visualization of the contents mapped at a continuous grid. We manually select six global coordinates and visualize the corresponding sampled normal features.

Comparison

Concept vs Real results





Mixed normal & abnormal

Separated normal & abnormal

Mixed normal & abnormal

Performance (MVTec AD)

Persp.	Method	#Entry	Detection	Localization
Local	VQ	64	97.0±0.85	96.0±0.06
		256	97.4±0.19	96.1±0.24
	Attention	64	95.9±0.36	96.4±0.06
		256	95.5±0.26	96.2±0.06
	GRAD	64	**98.0±0.05**	**97.1±0.07**
Global	VQ	16	78.9±1.28	86.7±1.93
		64	76.9±0.73	86.1±0.49
	Attention	16	79.8±1.22	88.0±0.27
		64	80.1±2.56	87.8±1.80
	GRAD	16	**93.0±0.34**	**95.7±0.04**

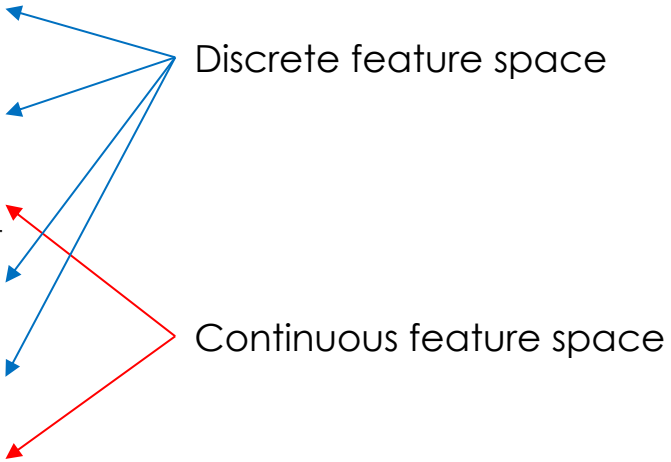


Table 1: Performance evaluation of the different feature space in a unified setting. #Entry denotes the number of features in each feature space and Persp. indicates the perspective (local or global).

Local	Global	Refine	Jitter	Detect	Localize
	✓			93.0	95.7
✓				98.0	97.1
✓	✓			98.9	97.7
✓	✓	✓		99.2	97.8
✓	✓	✓	✓	99.3	97.8

Table 5: Ablation studies on the proposed method in the unified setting using MVTec AD.

Detection Performance (MVTec AD)

Class	US [4]	PaDiM [8]	MKD [32]	DRAEM [41]	RD4AD [9]	PatchCore [30]	UniAD [40]	HVQ-T [23]	GRAD (Ours)
Bottle	84.0/99.0	97.9/99.9	98.7/99.4	97.5/99.2	98.7/100	100/100	99.7/100	100/-	100±0.00/100
Cable	60.0/86.2	70.9/92.7	78.2/89.2	57.8/91.8	85.0/95.0	99.7/99.4	95.2/97.6	99.0/-	99.5±0.15/99.8
Capsule	57.6/86.1	73.4/91.3	68.3/80.5	65.3/98.5	95.5/96.3	90.9/97.8	86.9/85.3	95.4/-	96.9±0.37/97.9
Hazelnut	95.8/93.1	85.5/92.0	97.1/98.4	93.7/100	87.1/99.9	100/100	99.8/99.9	100/-	99.9±0.09/100
Metal Nut	62.7/82.0	88.0/98.7	64.9/73.6	72.8/98.7	99.4/100	99.9/100	99.2/99.0	99.9/-	99.9±0.08/99.9
Pill	56.1/87.9	68.8/93.3	79.7/82.7	82.2/98.9	52.6/96.6	96.9/96.0	93.7/88.3	95.8/-	97.8±0.37/98.7
Screw	66.9/54.9	56.9/85.8	75.6/83.3	92/93.9	97.3/97.0	90.1/97.0	87.5/91.9	95.6/-	97.5±0.43/99.0
Toothbrush	57.8/95.3	95.3/96.1	75.3/92.2	90.6/100	99.4/99.5	100/99.7	94.2/95.0	93.6/-	99.3±0.13/97.2
Transistor	61.0/81.8	86.6/97.4	73.4/85.6	74.8/93.1	92.4/96.7	99.7/100	99.8/100	99.7/-	99.9±0.09/99.9
Zipper	78.6/91.9	79.7/90.3	87.4/93.2	98.8/100	99.6/98.5	94.7/99.5	95.8/96.7	97.9/-	99.2±0.15/99.2
Carpet	86.6/91.6	93.8/99.8	69.8/79.3	98.0/97.0	97.1/98.9	97.1/98.7	99.8/99.9	99.9/-	100±0.03/100
Grid	69.2/81.0	73.9/96.7	83.8/78.0	99.3/99.9	99.7/100	96.3/97.9	98.2/98.5	97.0/-	100±0.0/100
Leather	97.2/88.2	99.9/100	93.6/95.1	98.7/100	100/100	100/100	100/100	100/-	100±0.00/100
Tile	93.7/99.1	93.3/98.1	89.5/91.6	99.8/99.6	97.5/99.3	99.0/98.9	99.3/99.0	99.2/-	100±0.00/100
Wood	90.6/97.7	98.4/99.2	93.4/94.3	99.8/99.1	99.2/99.2	99.5/99.0	98.6/97.9	97.2/-	99.4±0.13/99.7
Mean	74.5/87.7	84.2/95.5	81.9/87.8	88.1/98.0	93.4/98.5	97.6/99.0	96.5/96.6	98.0/-	**99.3±0.14**/**99.4**

Table 2: Quantitative results for anomaly detection, evaluated with AUROC metric on MVTec-AD. All methods are evaluated under the unified and separate settings.

Localization Performance (MVTec AD)

Class	US [4]	PaDiM [8]	MKD [32]	DRAEM [41]	RD4AD [9]	PatchCore [30]	UniAD [40]	HVQ-T [23]	GRAD (Ours)
Bottle	67.9/97.8	96.1/98.2	91.8/96.3	87.6/99.1	97.7/98.7	98.4/98.6	98.1/98.1	98.3/-	98.4±0.11/98.5
Cable	78.3/91.9	81.0/96.7	89.3/82.4	71.3/94.7	83.1/97.4	96.7/98.5	97.3/96.8	98.1/-	98.3±0.16/98.3
Capsule	85.5/96.8	96.9/98.6	88.3/95.9	50.5/94.3	98.5/98.7	94.8/98.9	98.5/97.9	98.8/-	98.6±0.02/98.4
Hazelnut	93.7/98.2	96.3/98.1	91.2/94.6	96.9/99.7	98.7/98.9	98.6/98.7	98.1/98.8	98.8/-	98.7±0.19/98.7
Metal Nut	76.6/97.2	84.8/97.3	64.2/86.4	62.2/99.5	94.1/97.3	98.3/98.4	94.8/95.7	96.3/-	97.6±0.21/97.8
Pill	80.3/96.5	87.7/95.7	69.7/89.6	94.4/97.6	96.5/98.2	97.3/97.6	95.0/95.1	97.1/-	97.9±0.10/97.9
Screw	90.8/97.4	94.1/98.4	92.1/96.0	95.5/97.6	99.4/99.6	98.0/99.4	98.3/97.4	98.9/-	99.2±0.06/99.1
Toothbrush	86.9/97.9	95.6/98.8	88.9/96.1	97.7/98.1	99.0/99.1	98.4/98.7	98.4/97.8	98.6/-	98.8±0.03/94.1
Transistor	68.3/73.7	92.3/97.6	71.7/76.5	64.5/90.9	86.4/92.5	94.9/96.4	97.9/98.7	97.9/-	98.3±0.25/98.7
Zipper	84.2/95.6	94.8/98.4	86.1/93.9	98.3/98.8	98.1/98.2	95.8/98.9	96.8/96.0	97.5/-	97.9±0.07/97.7
Carpet	88.7/93.5	97.6/99.0	95.5/95.6	98.6/95.5	98.8/98.9	98.9/99.1	98.5/98.0	98.7/-	98.7±0.02/98.8
Grid	64.5/89.9	71.0/97.1	82.3/91.8	98.7/99.7	99.2/99.3	96.9/98.7	96.5/94.6	97.0/-	98.0±0.05/98.0
Leather	95.4/97.8	84.8/99.0	96.7/98.1	97.3/98.6	99.4/99.4	99.3/99.3	98.8/98.3	98.8/-	98.7±0.20/99.2
Tile	82.7/92.5	80.5/94.1	85.3/82.8	98.0/99.2	95.6/95.6	95.9/95.9	91.8/91.8	92.2/-	94.2±0.20/94.1
Wood	83.3/92.1	89.1/94.1	80.5/84.8	96.0/96.4	96.0/95.3	94.4/95.1	93.2/93.4	92.4/-	94.0±0.10/94.6
Mean	81.8/93.9	89.5/97.4	84.9/90.7	87.2/97.3	96.0/97.8	97.1/**98.1**	96.8/96.6	97.3/-	**97.8±0.12**/97.9

Table 3: Quantitative results for anomaly localization, evaluated with AUROC metric on MVTec-AD. All methods are evaluated under the unified and separate settings.

Qualitative Results (MVTec AD)

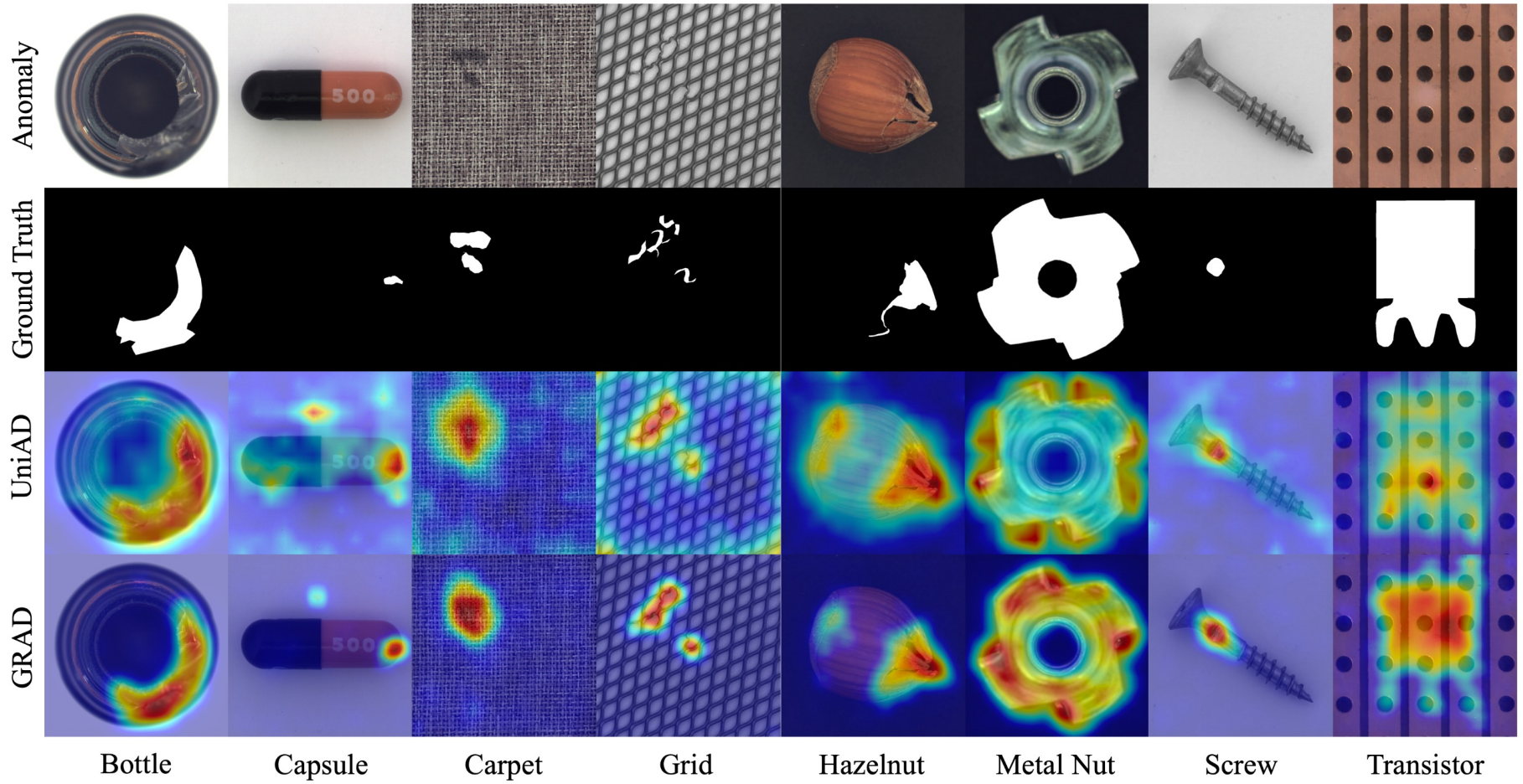


Fig. 5: Qualitative results of GRAD on MVTec AD. Each row of the figure represents anomaly images, corresponding ground truths, results from UniAD, and our results.

Performance (VisA)

Class		Detection				Localization			
		UniAD [40]	PatchCore [30]	OmniAL [42]	GRAD (ours)	UniAD [40]	PatchCore [30]	OmniAL [42]	GRAD (ours)
Complex Structure	PCB1	94.8/90.2	97.6/98.5	77.7/96.6	98.0/95.3	99.3/99.2	99.7/99.8	97.6/98.7	99.5/99.3
	PCB2	92.5/84.2	96.7/97.2	81.0/99.4	90.5/93.3	97.6/96.5	98.0/98.7	93.9/83.2	97.3/96.6
	PCB3	86.6/90.7	97.3/98.5	88.1/96.9	96.2/96.5	98.1/98.0	99.3/99.4	94.7/98.4	98.7/98.5
	PCB4	99.3/97.4	99.7/99.7	95.3/97.4	99.1/98.1	97.6/97.2	97.7/98.2	97.1/98.5	98.3/98.0
Multiple Instances	Macaroni1	90.4/90.2	94.7/97.4	92.6/96.9	93.5/95.2	99.1/99.0	99.0/99.7	98.6/98.9	99.0/99.0
	Macaroni2	82.8/77.4	78.6/76.7	75.2/89.9	95.9/88.8	97.7/97.4	96.1/98.6	97.9/99.1	98.8/98.8
	Capsules	70.7/80.3	75.0/76.3	90.6/87.9	87.3/93.5	98.1/98.5	99.1/99.2	99.4/98.6	99.4/99.6
	Candle	97.0/90.2	94.7/99.4	86.8/85.1	95.9/96.9	99.1/99.0	98.3/99.3	95.8/90.5	99.3/99.2
Single Instance	Cashew	93.8/92.9	97.3/97.8	88.6/97.1	93.6/96.4	98.9/99.2	98.1/98.7	95.0/98.9	97.8/97.5
	Chewinggum	99.3/98.3	98.5/98.8	96.4/94.9	94.6/99.6	99.1/98.5	98.9/98.9	99.0/98.7	98.3/98.6
	Fryum	88.8/84.4	95.4/96.0	94.6/97.0	99.5/93.2	97.7/96.7	89.8/92.4	92.1/89.3	96.7/96.1
	Pipe fryum	97.0/91.8	99.2/99.8	86.1/91.4	98.0/98.8	99.3/99.3	97.5/98.9	98.2/99.1	99.4/98.9
Mean		91.1/89.0	93.7/94.7	87.8/94.2	**95.2**/**95.5**	98.5/98.2	97.5/98.5	96.6/96.0	98.5/98.3

Table 4: Quantitative results for anomaly detection and localization, evaluated on VisA. All methods are evaluated under the unified and separate settings.

Qualitative Results (VisA)

Not provided

Appendix A

Cited anomaly detection models

US [4]
Uninformed Student

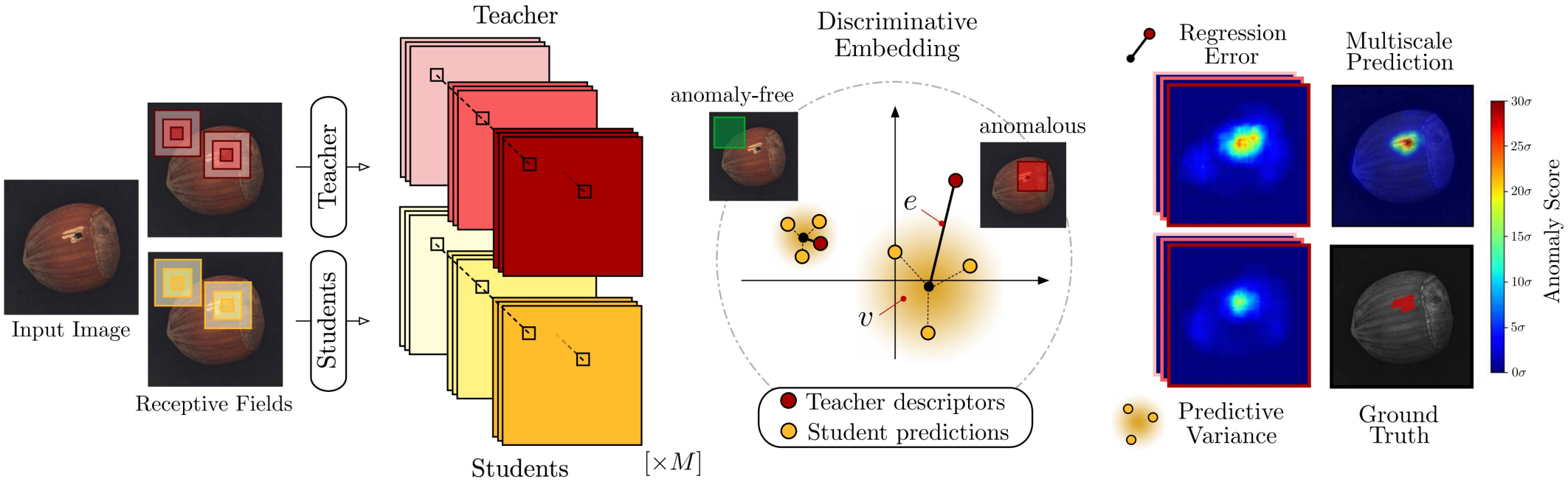


Figure 2: Schematic overview of our approach. Input images are fed through a teacher network that densely extracts features for local image regions. An ensemble of M student networks is trained to regress the output of the teacher on anomaly-free data. During inference, the students will yield increased regression errors e and predictive uncertainties v in pixels for which the receptive field covers anomalous regions. Anomaly maps generated with different receptive fields can be combined for anomaly segmentation at multiple scales.

- **Teacher**: pre-trained model by triplet learning with large dataset
- **Student**: trained to predict feature map of teacher model

PaDiM [8]

Patch Distribution Modeling

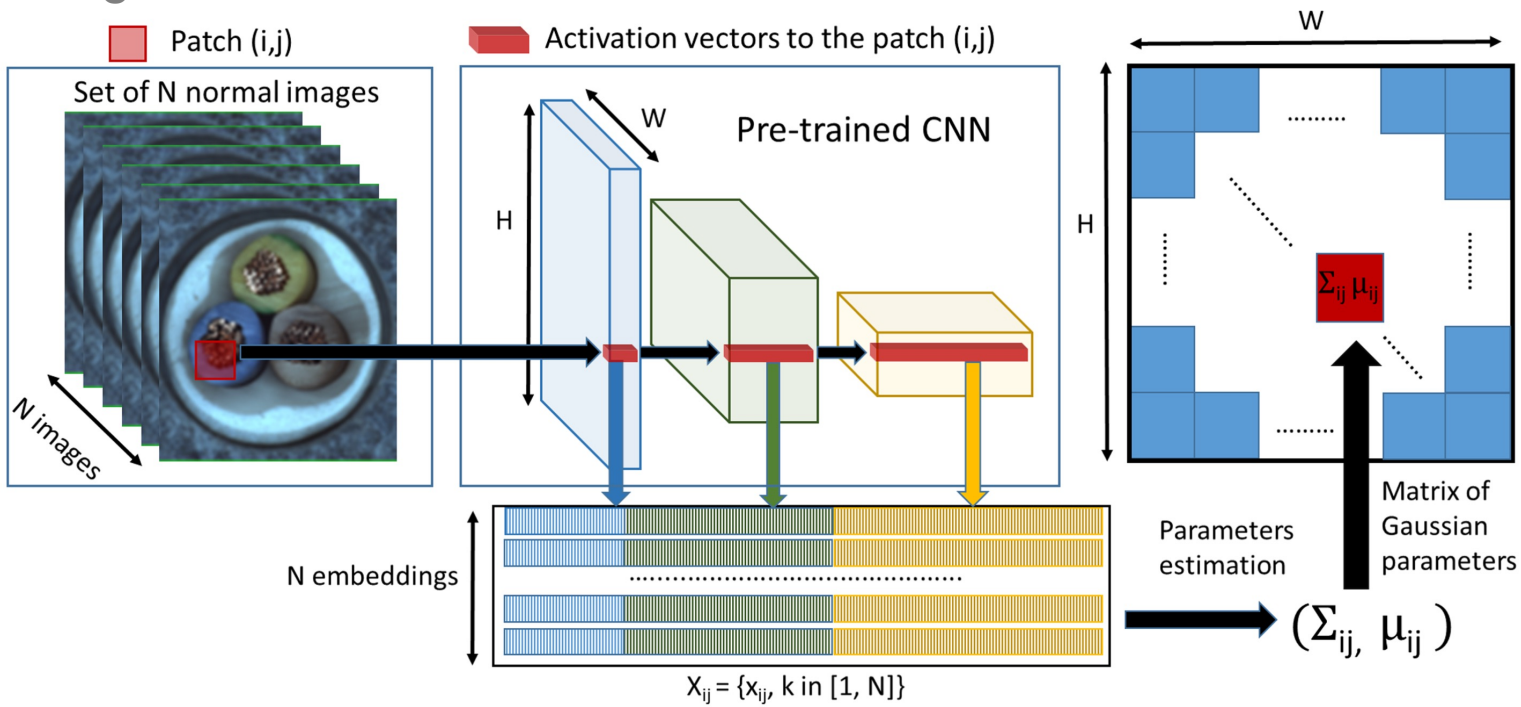


Fig. 2. For each image patch corresponding to position (i, j) in the largest CNN feature map, PaDiM learns the Gaussian parameters (μ_{ij}, Σ_{ij}) from the set of N training embedding vectors $X_{ij} = \{x_{ij}^k, k \in [\![1, N]\!]\}$, computed from N different training images and three different pretrained CNN layers. (Color figure online)

- **Training**: Memorizing patch-wise normal Gaussian distribution of normal samples
- **Inference**: Measure Mahalanobis between input feature and mean vector

MKD [32]

Multiresolution Knowledge Distillation

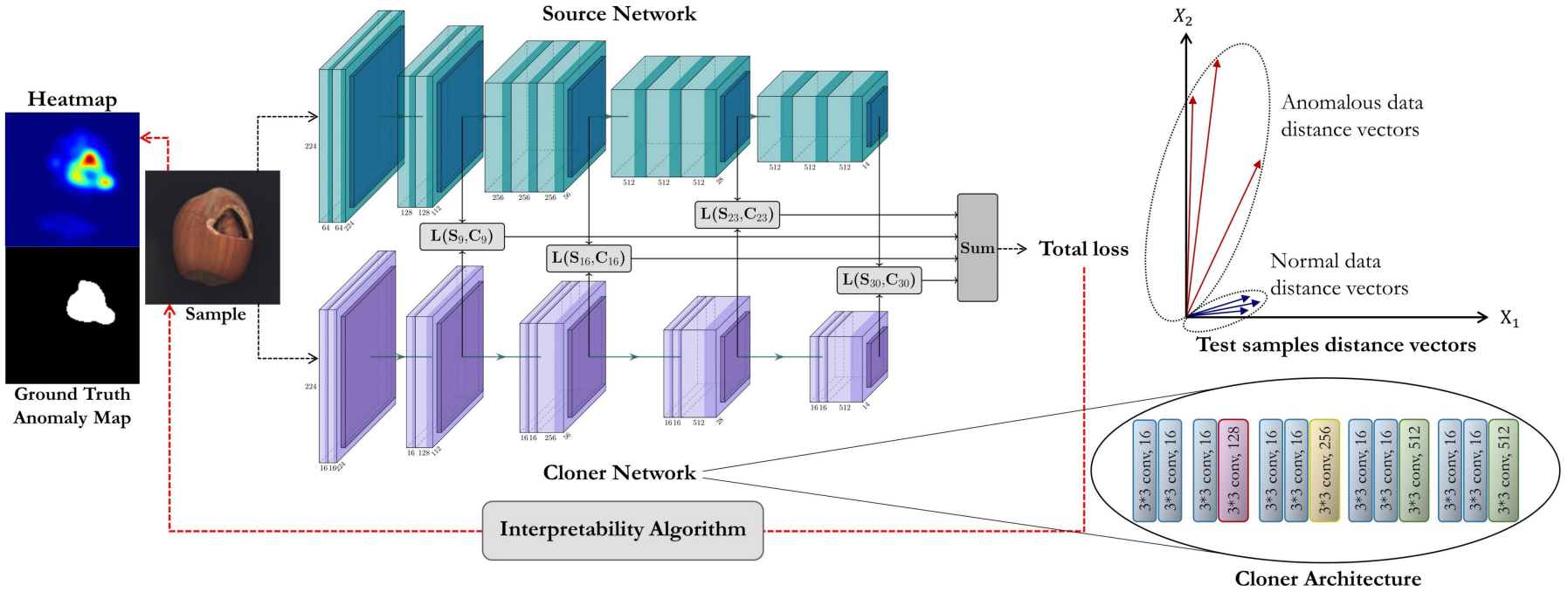


Figure 2: Visualized summary of our proposed framework. A smaller cloner network, C, is trained to imitate the *whole* behavior of a source network, S (VGG-16), on normal data. The discrepancy of their intermediate behavior is formulated by a total loss function and is used to detect anomalies at the test time. A hypothetical example of distance vectors between the activations of C and S on anomalous and normal data is also depicted. Interpretability algorithms are employed to yield pixel-precise anomaly localization maps.

- **Source Net**: pre-trained model by large dataset (ImageNet)
- **Cloner Net**: trained to predict feature map of source network

DRAEM [41]

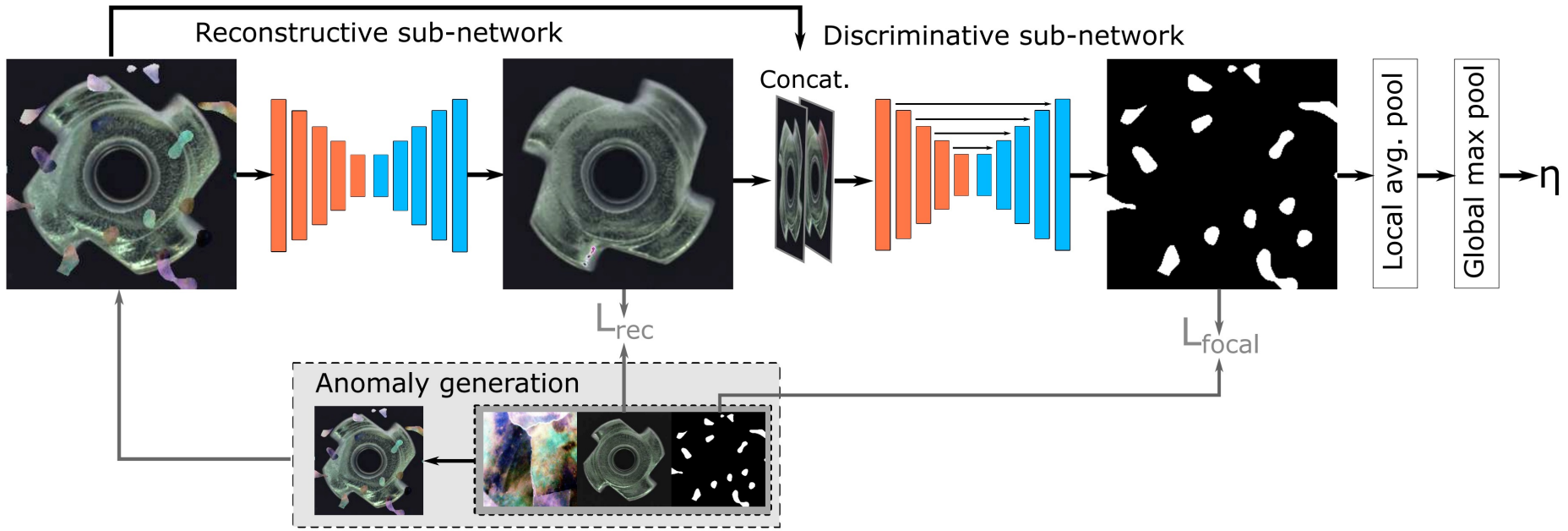


Figure 3. The anomaly detection process of the proposed method. First anomalous regions are implicitly detected and inpainted by the reconstructive sub-network trained using L_{rec}. The output of the reconstructive sub-network and the input image are then concatenated and fed into the discriminative sub-network. The segmentation network, trained using the Focal loss L_{focal}[14], localizes the anomalous region and produces an anomaly map. The image level anomaly score η is acquired from the anomaly score map.

End-to-End defect segmentation & anomaly scoring

RD4AD [9]

Reverse Distillation for Anomaly Detection

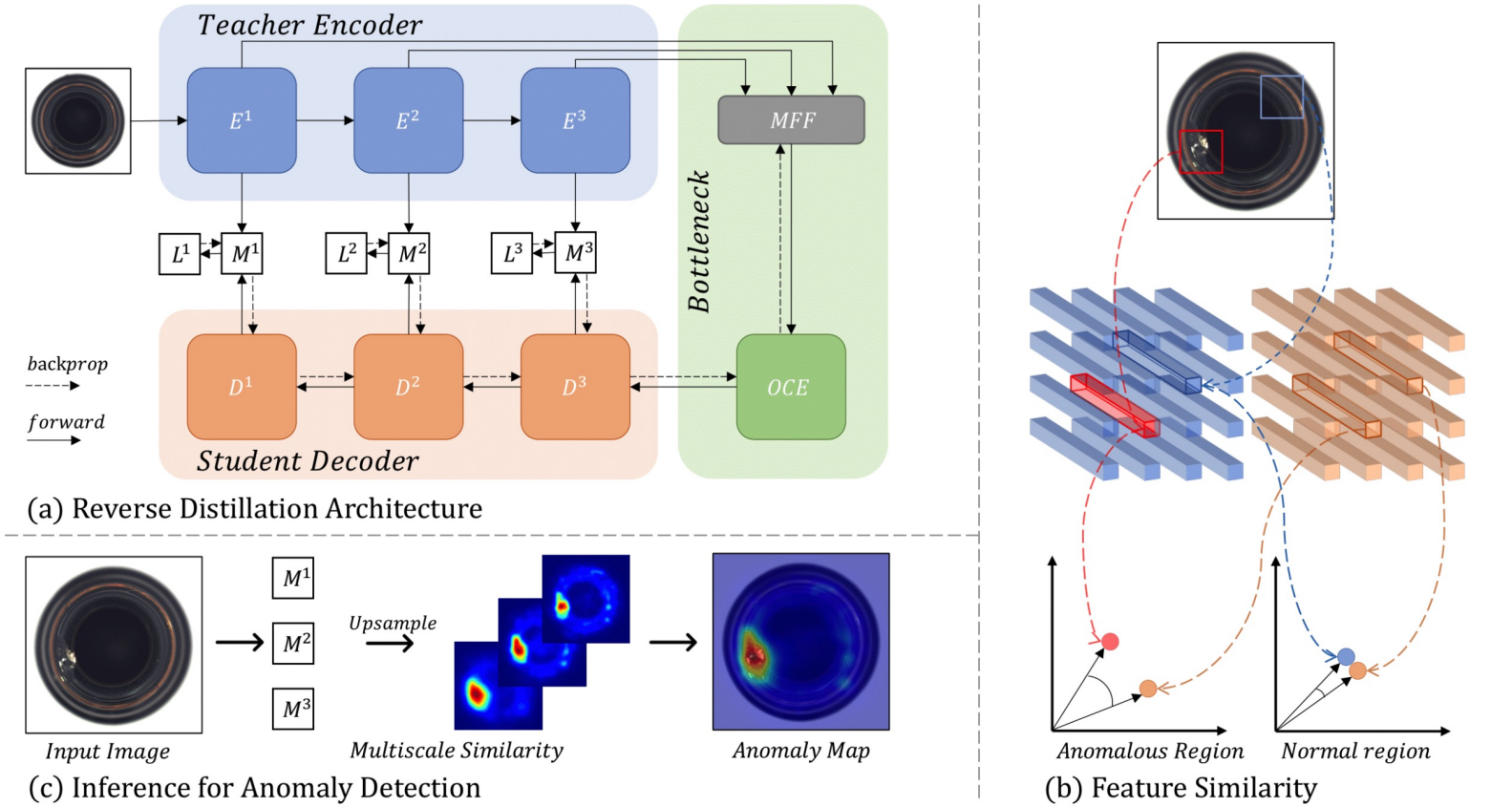


Figure 3. Overview of our reverse distillation framework for anomaly detection and localization. (a) Our model consists of a pre-trained teacher encoder E, a trainable one-class bottleneck embedding module (OCBE), and a student decoder D. We use a multi-scale feature fusion (MFF) block to ensemble low- and high-level features from E and map them onto a compact code by one-class embedding (OCE) block. During training, the student D learns to mimic the behavior of E by minimizing the similarity loss $\mathcal{L}$. (b) During inference, E extracts the features truthfully, while D outputs anomaly-free ones. A low similarity between the feature vectors at the corresponding position of E and D implies an abnormality. (c) The final prediction is calculated by the accumulation of multi-scale similarity maps M.

Karsten Roth, et al. "Towards total recall in industrial anomaly detection." CVPR. 2022.

PatchCore [30]

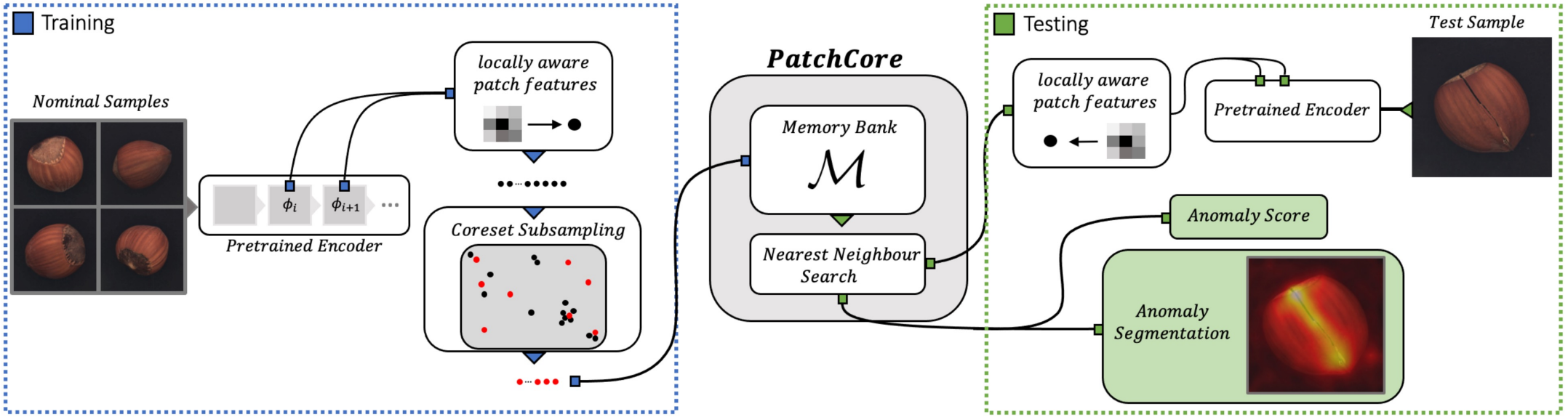


Figure 2. Overview of *PatchCore*. Nominal samples are broken down into a memory bank of neighbourhood-aware patch-level features. For reduced redundancy and inference time, this memory bank is downsampled via greedy coreset subsampling. At test time, images are classified as anomalies if at least one patch is anomalous, and pixel-level anomaly segmentation is generated by scoring each patch-feature.

UniAD [40]

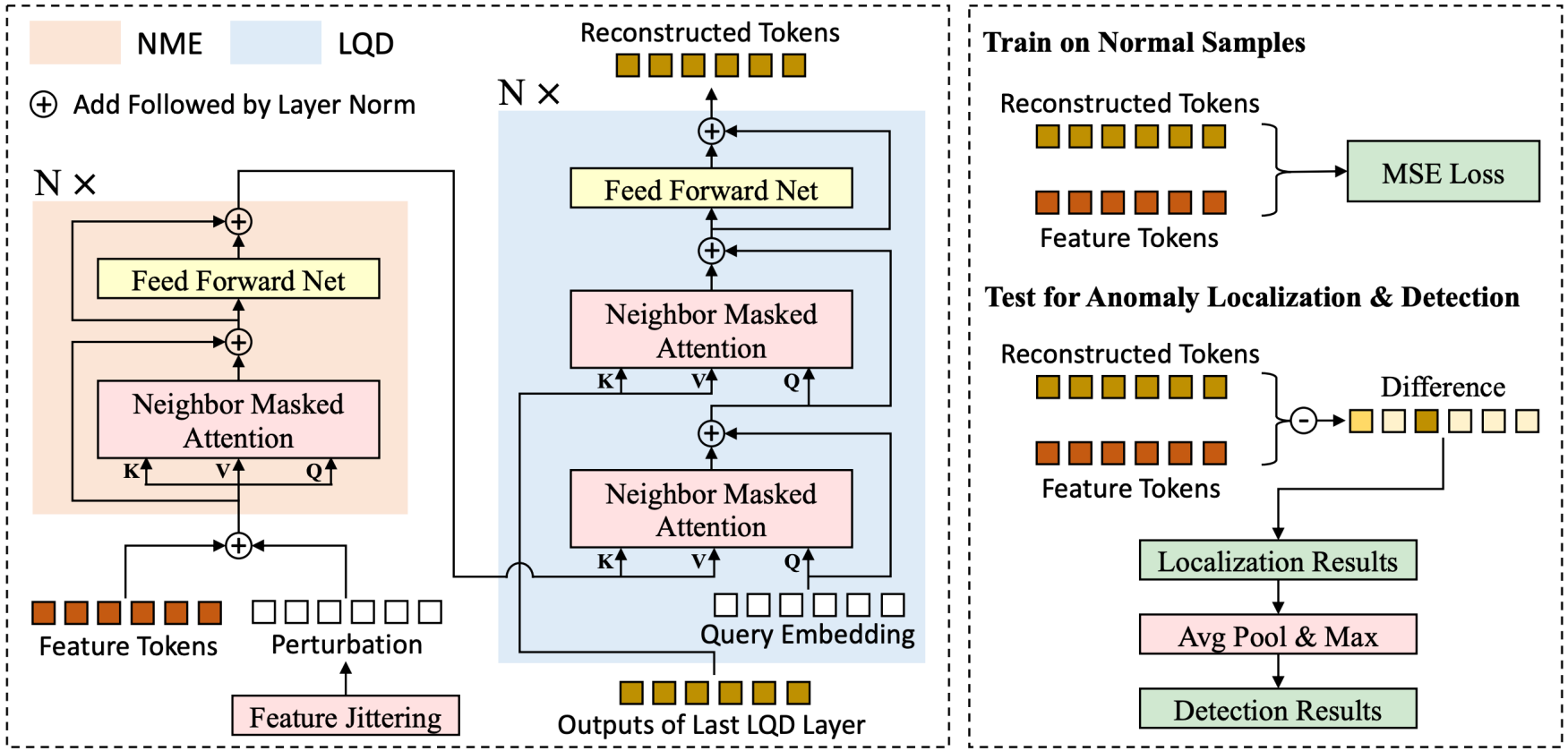


Figure 3: **Framework** of UniAD, consisting of a Neighbor Masked Encoder (NME) and a Layer-wise Query Decoder (LQD). Each layer in LQD employs a *learnable query embedding* to help model the complex training data distribution. The full attention in transformer is replaced by *neighbor masked attention* to avoid the information leak from the input to the output. The *feature jittering* strategy encourages the model to recover the correct message with noisy inputs. All the three improvements assist the model against learning the "identical shortcut" (see Sec. 3.1 and Fig. 2 for details).

HVQ-T [23]

Hierarchical Vector Quantized Transformer

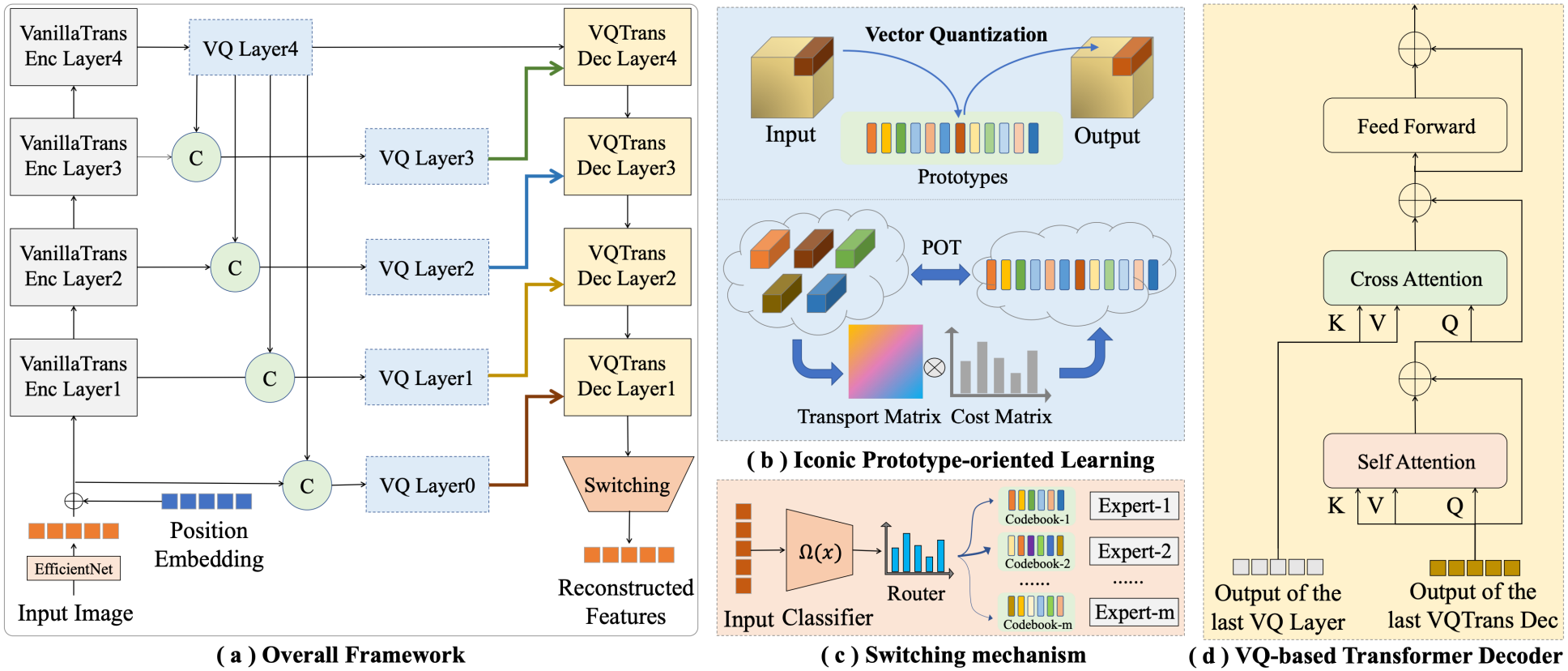


Figure 2: (a) The overall framework of our HVQ-Trans. (b) Each VQ-based Layer replaces continuous features with iconic prototypes, equipped with the POT module to promote better learning and scoring. (c) The codebook and expert network are switched for individual image. (d) The detailed structure of each VQ-based Transformer decoder, where the prototypes are integrated via cross-attention.

Appendix B

How to speed up paperwork

Sharing

Table 1: Anomaly detection/localization results with AUROC metric on MVTec-AD. All methods are evaluated under the one-for-all settings. The learned model is applied to detect anomalies for all categories without fine-tuning. The best results are bold with black.

	Category	US[46]	PSVDD[47]	PaDiM[48]	MKD[49]	DRAEM[50]	SimpleNet[51]	PatchCore[8]	RD4AD[52]	UTRAD[33]	UniAD[10]	Ours
Object	Bottle	84.0 / 67.9	85.5 / 86.7	97.9 / 96.1	98.7 / 91.8	97.5 / 87.6	97.7 / 91.2	**100** / 97.4	98.7 / 97.7	**100** / 96.4	99.7 / 98.1	**100** ± 0.00 / **98.3** ± 0.04
	Cable	60.0 / 78.3	64.4 / 62.2	70.9 / 81.0	78.2 / 89.3	57.8 / 71.3	87.6 / 88.1	95.3 / 93.6	85.0 / 83.1	97.8 / 97.1	95.2 / 97.3	**99.0** ± 0.29 / **98.1** ± 0.04
	Capsule	57.6 / 85.5	61.3 / 83.1	73.4 / 96.9	68.3 / 88.3	65.3 / 50.5	78.3 / 89.7	**96.8** / 98.0	95.5 / 98.5	82.0 / 97.2	86.9 / 98.5	95.4 ± 0.21 / **98.8** ± 0.01
	Hazelnut	95.8 / 93.7	83.9 /97.4	85.5 / 96.3	97.1 / 91.2	93.7 / 96.9	99.2 / 95.7	99.3 / 97.6	87.1 / 98.7	99.8 / 98.2	99.8 / 98.1	**100** ± 0.08 / **98.8** ± 0.02
	Metal Nut	62.7 /76.6	80.9 / 96.0	88.0 / 84.8	64.9 / 64.2	72.8 / 62.2	85.1 / 90.9	99.1 / 96.3	99.4 / 94.1	94.7 / **96.4**	99.2 / 94.8	**99.9** ± 0.02 / 96.3 ± 0.09
	Pill	56.1 / 80.3	89.4 / 96.5	68.8 / 87.7	79.7 / 69.7	82.2 / 94.4	78.3 / 89.7	86.4 / 90.8	52.6 / 96.5	89.7 / 95.7	93.7 / 95.0	**95.8** ± 0.49 / **97.1** ± 0.04
	Screw	66.9 / 90.8	80.9 / 74.3	56.9 / 94.1	75.6 / 92.1	92.0 / 95.5	45.5 / 93.7	94.2 / 98.9	**97.3** / **99.4**	75.1 / 95.2	87.5 / 98.3	95.6 ± 0.60 / 98.9 ± 0.05
	Toothbrush	57.8 / 86.9	99.4 / 98.0	95.3 / 95.6	75.3 / 88.9	90.6 / 97.7	94.7 / 97.5	**100** / 98.8	99.4 / **99.0**	89.7 / 97.5	94.2 / 98.4	93.6 ± 0.66 / 98.6 ± 0.04
	Transistor	61.0 / 68.3	77.5 / 78.5	86.6 / 92.3	73.4 / 71.7	74.8 / 64.5	82.0 / 86.0	98.9 / 92.3	92.4 / 86.4	92.0 / 91.5	**99.8** / **97.9**	99.7 ± 0.12 / **97.9** ± 0.04
	Zipper	78.6 / 84.2	77.8 / 95.1	79.7 / 94.8	87.4 / 86.1	98.8 / **98.3**	99.1 / 97.0	97.1 / 95.7	**99.6** / 98.1	95.5 / 97.3	95.8 / 96.8	97.9 ± 0.15 / 97.5 ± 0.10
Texture	Carpet	86.6 / 88.7	63.3 / 78.6	93.8 / 97.6	69.8 / 95.5	98.0 / 98.6	95.9 / 92.4	97.0 / 98.1	97.1 / **98.8**	80.3 / 94.4	99.8 / 98.5	**99.9** ± 0.03 / 98.7 ± 0.03
	Grid	69.2 / 64.5	66.0 / 70.8	73.9 / 71.0	83.8 / 82.3	99.3 / 98.7	49.8 / 46.7	91.4 / 98.4	**99.7** / **99.2**	93.9 / 95.2	98.2 / 96.5	97.0 ± 0.69 / 97.0 ± 0.06
	Leather	97.2 / 95.4	60.8 / 93.5	99.9 / 84.8	93.6 / 96.7	98.7 / 97.3	93.9 / 96.9	**100** / 99.2	**100** / **99.4**	99.8 / 98.4	**100** / 98.8	**100** ± 0.00 / 98.8 ± 0.02
	Tile	93.7 / 82.7	88.3 / 92.1	93.3 / 80.5	89.5 / 85.3	**99.8** / **98.0**	93.7 / 93.1	96.0 / 90.3	97.5 / 95.6	98.8 / 94.2	99.3 / 91.8	99.2 ± 0.32 / 92.2 ± 0.42
	Wood	90.6 / 83.3	72.1 / 80.7	98.4 / 89.1	93.4 / 80.5	**99.8** / **96.0**	95.2 / 84.8	93.8 / 90.8	99.2 / 96.0	99.7 / 89.4	98.6 / 93.2	97.2 ± 0.40 / 92.4 ± 0.16
	Mean	74.5 / 81.8	76.8 / 85.6	84.2 / 89.5	81.9 / 84.9	88.1 / 87.2	85.1 / 88.9	96.4 / 95.7	93.4 / 96.0	92.6 / 95.6	96.5 / 96.8	**98.0** ± 0.11 / **97.3** ± 0.05

Table Sharing

Code Sharing

Class	US [4]	PaDiM [8]	MKD [32]	DRAEM [41]	RD4AD [9]	PatchCore [30]	UniAD [40]	HVQ-T [23]	GRAD (Ours)
Bottle	84.0/99.0	97.9/99.9	98.7/99.4	97.5/99.2	98.7/100	100/100	99.7/100	100/-	100±0.00/100
Cable	60.0/86.2	70.9/92.7	78.2/89.2	57.8/91.8	85.0/95.0	99.7/99.4	95.2/97.6	99.0/-	99.5±0.15/99.8
Capsule	57.6/86.1	73.4/91.3	68.3/80.5	65.3/98.5	95.5/96.3	90.9/97.8	86.9/85.3	95.4/-	96.9±0.37/97.9
Hazelnut	95.8/93.1	85.5/92.0	97.1/98.4	93.7/100	87.1/99.9	100/100	99.8/99.9	100/-	99.9±0.09/100
Metal Nut	62.7/82.0	88.0/98.7	64.9/73.6	72.8/98.7	99.4/100	99.9/100	99.2/99.0	99.9/-	99.9±0.08/99.9
Pill	56.1/87.9	68.8/93.3	79.7/82.7	82.2/98.9	52.6/96.6	96.9/96.0	93.7/88.3	95.8/-	97.8±0.37/98.7
Screw	66.9/54.9	56.9/85.8	75.6/83.3	92/93.9	97.3/97.0	90.1/97.0	87.5/91.9	95.6/-	97.5±0.43/99.0
Toothbrush	57.8/95.3	95.3/96.1	75.3/92.2	90.6/100	99.4/99.5	100/99.7	94.2/95.0	93.6/-	99.3±0.13/97.2
Transistor	61.0/81.8	86.6/97.4	73.4/85.6	74.8/93.1	92.4/96.7	99.7/100	99.8/100	99.7/-	99.9±0.09/99.9
Zipper	78.6/91.9	79.7/90.3	87.4/93.2	98.8/100	99.6/98.5	94.7/99.5	95.8/96.7	97.9/-	99.2±0.15/99.2
Carpet	86.6/91.6	93.8/99.8	69.8/79.3	98.0/97.0	97.1/98.9	97.1/98.7	99.8/99.9	99.9/-	100±0.03/100
Grid	69.2/81.0	73.9/96.7	83.8/78.0	99.3/99.9	99.7/100	96.3/97.9	98.2/98.5	97.0/-	100±0.0/100
Leather	97.2/88.2	99.9/100	93.6/95.1	98.7/100	100/100	100/100	100/100	100/-	100±0.00/100
Tile	93.7/99.1	93.3/98.1	89.5/91.6	99.8/99.6	97.5/99.3	99.0/98.9	99.3/99.0	99.2/-	100±0.00/100
Wood	90.6/97.7	98.4/99.2	93.4/94.3	99.8/99.1	99.2/99.2	99.5/99.0	98.6/97.9	97.2/-	99.4±0.13/99.7
Mean	74.5/87.7	84.2/95.5	81.9/87.8	88.1/98.0	93.4/98.5	97.6/99.0	96.5/96.6	98.0/-	**99.3±0.14**/**99.4**

Table 2: Quantitative results for anomaly detection, evaluated with AUROC metric on MVTec-AD. All methods are evaluated under the unified and separate settings.