SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 6, 2026

MASF: MASKED ASYMMETRIC SPECTRAL FLOW FOR UNSUPERVISED INDUSTRIAL ANOMALY DETECTION

Maheem Khowaja[*1], **Dr. Shahid Khan Yousafzai**[2]

[*1,2]*Department of Artificial Intelligence and Robotics*
SZABIST Karachi, Pakistan

DOI: https://doi.org/10.5281/zenodo.20588258

Keywords
anomaly detection, frequency domain, spectral decomposition, masked feature distillation, MVTec AD, industrial inspection, unsupervised learning, computer vision.

Article History
Received: 07 April 2026
Accepted: 19 May 2026
Published: 08 June 2026

Corresponding Author: *
Maheem Khowaja

Abstract
An important problem in the industrial quality control application is that of unsupervised anomaly detection, in which only defect free training images are available. All the currently available state-of-the-art techniques such as PaDiM, PatchCore, RD++, CFA, ISSTAD, and ADTR are in the spatial domain and miss the part of the information in the frequency domain of CNN feature maps that has anomaly-discriminative information. We introduce the first framework to leverage frequency domain at the CNN feature level, called MASF (Masked Asymmetric Spectral Flow). MASF introduces five novel components: (1) a Spectral Frequency Decomposition Module (SDM) based on 2D FFT on intermediate feature maps; (2) Asymmetric Masked Feature Distillation (AMFD) using dual spatial-frequency domain masking and Spectral-Spatial Cross-Attention (SSCA) fusion; (3) a Spectral-Anchored Memory Bank (SAMB) for rotation-robust prototype retrieval; (4) Uncertainty-Gated Hierarchical Score Fusion (UGHF) with learnable per-scale precision weights; and (5) Test-Time Spectral Augmentation (TTSA) by FFT phase perturbation. On the Bottle category, evaluated on the MVTec Anomaly Detection benchmark, MASF gets Image-AUROC = 99.9999, Pixel-AUROC = 0.9853, PRO = 0.9461, and AP = 99.9999. In 11 stable training categories, MASF reaches the performance of mean Image-AUROC = 0.8008 and mean Pixel-AUROC = 0.9380 with just 15 training epochs. The design of MASF is directly motivated by the results of FFT spectral analysis which shows that the frequency signature of the normal image is different from that of the anomalous image, as shown by industrial images.

1. INTRODUCTION

Industrial manufacturing requires zero-defect quality control, but it is very costly to annotate all the possible defect types: they are rare, varied, and occur at random. This is called unsupervised anomaly detection (UAD) and it works by training only on normal, defect-free images, and detect any deviations in the test images. The new dataset MVTec Anomaly Detection (MVTec AD) [1] including 15 industrial categories of high-resolution images and pixel-accurate anomaly masks has proven to be the most popular benchmark and has accelerated the methodological development in the field.

1.1 Limitations of spatial domain methods.

Near-perfect image-level detection is achieved with recent UAD methods, such as: PaDiM [2] extracts the patch distributions with multivariate Gaussians, PatchCore [3] constructs a coreset-subsampled feature memory bank (99.6% mean Image-AUROC), RD++ [4] exploits the reverse

distillation with optimal transport, CFA [5] adapts features through coupled hyperspheres, and ISSTAD [6] utilizes Masked Autoencoders for incremental self-supervised learning, while ADTR [7] reconstructs features using Vision Transformers. All of them are operated in the spatial domain only. This is a basic limitation: Texture anomalies occur over a range of frequencies that are periodic; micro-crack information is encoded at high frequencies; structural deformations are encoded at low frequencies. Models with blind frequency domain feature content should learn the frequency domain without a label.

1.2 Proposed approach. In this paper, we propose the first UAD framework that explicitly decomposes CNN feature maps in the frequency domain, which is solved by decomposing the CNN feature maps into a sum of orthogonal sub-models. Section 4.7 shows there are measurable differences between the FFT magnitude spectra of normal and anomalous images from the industrial field, giving this approach a strong motivating signal. MASF introduces five new technical contributions, each dealing with a particular limitation of previous approaches.

1.3 Contributions

a. Spectral Frequency Decomposition Module (SDM): First 2D FFT decomposition of CNN feature maps for anomaly detection, complementary low frequency/structural defects anomaly channel and high frequency/texture defects anomaly channel.

b. Spatial patch masking (35%) and frequency coefficient masking (25%): Simultaneous patch masking in the pixel domain and simultaneous masking in the frequency domain with the fusion of the two domains by the Spectral-Spatial Cross-Attention (SSCA) module — extending pixel-space masking of ISSTAD/MAE into the frequency domain.

c. Further prototypes are stored in a joint space-spectral memory bank (SAMB) that supports rotation robustness by virtue of the FFT magnitude shift-invariance without explicit spatial registration (cf. FR-PatchCore [13]).

d. Uncertainty-Gated Hierarchical Fusion (UGHF): A learnable per-scale log-variance parameter automatically down-weights unreliable scales — instead of fixed multi-scale fusion in PaDiM, RD++, and ADTR.

e. Test-Time Spectral Augmentation (TTSA): FFT phase perturbation at test-time offers orthogonally diverse views with zero training cost — a first test-time augmentation in anomaly detection for frequency domain.

2. Related Work

2.1 Reconstruction-Based Anomaly Detection

The early UAD methods involve training an autoencoder or variational autoencoder (VAE) on normal data to identify anomalies as high reconstruction error [8,9]. Variants of GANs [10] and attention-guided decoders [11] enhance spatial accuracy. There are some basic problems: strong generative models can reconstruct anomalies well [12] thus collapsing the error signal. To mitigate this, Memory-augmented autoencoders (AA) [12] employ prototype replacement but require extra hyper parameters. Unlike the usual approach to learning reconstruction spaces, MASF works directly in the teacher's pre-trained feature space, which naturally has more discriminative representations.

2.2 Incorporating Similarity and Memory Bank Methods

PaDiM [2] learns per-position multivariate Gaussians for CNN patch embeddings. PatchCore [3] greedily constructs a compact memory bank at a linear search cost with a good recall. CFA [5] fine-tunes patch descriptors using coupled hypersphere metric learning to minimize the gap between the ImageNet and industrial distribution. For this reason, the authors of FR-PatchCore [13] have introduced a technique for aligning PatchCore, called feature registration. MASF extends memory bank paradigm, where prototypes are stored in a shared space of both space and spectrum. The FFT magnitude spectrum is shift-variant, without the need of registration overhead, and is rotation robust.

2.3 Knowledge Distillation Methods

Student-teacher frameworks [14] take advantage of the difference between a pretrained frozen teacher model and a student model trained only on normal data. Reverse Distillation (RD) [15] is a technique that is based on the use of a decoder student with reversed data flow to limit over-generalization. RD++ [4] includes optimal transport feature regularization and simplex noise. RD-RE [16] proposes a cross-stage feature fusion approach using locally aware dynamic attention which obtains 99.70% Image-AUROC mean. MASF incorporates dual-domain masked training by enabling a student to reconstruct simultaneously from multiple views of different distortions, such as and including spatially masked and frequency-masked feature views, which are fused together using cross-attention, resulting in a richer anomaly signal than the spatial-only distillation.

2.4 Transformers or Masked Auto Encoder Methods

ISSTAD [6] uses MAE-style mask in pixel space and is trained incrementally in two stages. To overcome the identical mapping problem posed by attention, ADTR [7] proposes to reconstruct features that are pretrained. MASF is complementary: It is orthogonal to spatial masking and is a fundamentally different input modality.

2.5 The concept of frequency Domain Analysis in Vision.

Research gap. Most of the works involving Fourier-based methods are related to domain adaptation (FDA [17] and FACT [18]) and GAN artifact analysis [19] with input images. UAD has been studied relatively little in terms of frequency domain analysis. To the best of our knowledge, the concept of spectral decomposition of intermediate CNN feature maps has not been explored in any previous UAD research. MASF aims to address this gap by providing a comprehensive pipeline from feature decomposition to masked distillation, from memory storage to score fusion, and finally from test time augmentation.

3. Methodology

3.1 Problem Formulation

Let $X = \{x_1, \ldots, x_n\}$ be the training set of normal images from a single class. The goals are: (i) an image-level anomaly score $s(x) \in \mathbb{R}$, and (ii) a pixel-wise anomaly map $M(x) \in \mathbb{R}^{R^*H}$. The teacher network T (WideResNet-50 [21], pretrained on ImageNet, all parameters frozen) extracts feature maps $T\_l(x) \in \mathbb{R}^{B^*C\_L^*H\_L^*W\_L}$ at layers' $l \in \{1,2,3\}$, with C = {256, 512, 1024}. Only the student decoder modules, spectral projection layers, and the fusion gate are trained.

3.2 Architecture Overview

Figure 1 shows the MASF architecture. Four trainable modules attach to the frozen teacher: Spectral Decomposition Modules (SDM), Asymmetric Masked Student decoders with SSCA, a Spectral-Anchored Memory Bank, and an Uncertainty-Gated Fusion head. At inference, anomaly maps from three scales are fused and optionally refined with TTSA.

Figure 1. MASF architecture.

3.3 Spectral Frequency Decomposition Module (SDM)

Given teacher feature map $F \in \mathbb{R}^{B*C*H*W}$, SDM computes the 2D Discrete Fourier Transform (DFT) with orthonormal normalization: $\hat{F} = FFT_2(F, norm = 'ortho')$. A Boolean low-frequency mask M_low retains the central $r = 0.3$ fraction of the spectrum. The decomposition yields:

$F_{low} = IFFT_2(F \odot M_{low}), \quad F_{low} = IFFT_2(\hat{F} \odot \neg M_{low})$

A learnable spectral attention module (two-layer 1×1 convolution with GELU activation) produces channel-wise sigmoid gates that weight the contribution of *F_low* and *F_high*. A final 1×1 convolution + Batch Norm + GELU yields $F\_spec \in \mathbb{R}^{B*C*H*W}$. Unlike FDA [17] and FACT [18] which apply FFT to input images, SDM operates on *semantic feature maps* capturing defect-discriminative frequency content unavailable at pixel level.

3.4 Asymmetric Masked Feature Distillation (AMFD)

The student receives two complementary masked views of F_spec and reconstructs the teacher features — the asymmetry generates the anomaly signal. F_smask: 35% of 2×2 spatial patches of F_spec is set to zero, so that there is no direct copy of features. The coefficients of rfft2 (randomly taken) are zeroed so that the frequencies are not copied.

Cross-attention between views: Q = F_smask, K = V = F_fmask (4 attention heads). The output goes through two Residual Blocks and a linear projection. The training goal is designed to be a spatial cosine distillation and spectral consistency:

$$l = l_{cosine}\left(T(x), s\left(F_{Spec}\right)\right) + 0.5 * l_{MSE}\left(\left|FFT_2\left(S\left(F_{spec}\right)\right)\right|, \left|FFT_2(T(x))\right|\right)$$

3.5 Spectral-Anchored Memory Bank (SAMB)

For every patch embedding p that is a vector in ℝ^C from normal training images, we calculate: p_spec = MLP_2(pad_to_C(|FFT_1(p)|)). The joint prototype is equal to L2_norm (p + p_spec). There are also joint space runs of Greedy k-center core subsampling [3] (ratio 0.1, max 3,000/scale). When inference: s_bank = mean_k dist(p_joint(x), M). In contrast to explicit registration (cf. FR-PatchCore [13]), FFT magnitude shift-invariance gives implicit rotation robustness.

3.6 Uncertainty-Gated Hierarchical Fusion (UGHF)

The learned log-variance parameters {log σ^2_l} are fused together using three per-scale maps, {s_1, s_2, s_3} = 0.6× cosine score + 0.4× bank score.

$$s_{fused} = \sum_l \frac{exp\ exp\left(-log\sigma_l^2\right)}{Z} s_l,$$

$$Z = \sum_l exp\ exp\left(-log\sigma_l^2\right)$$

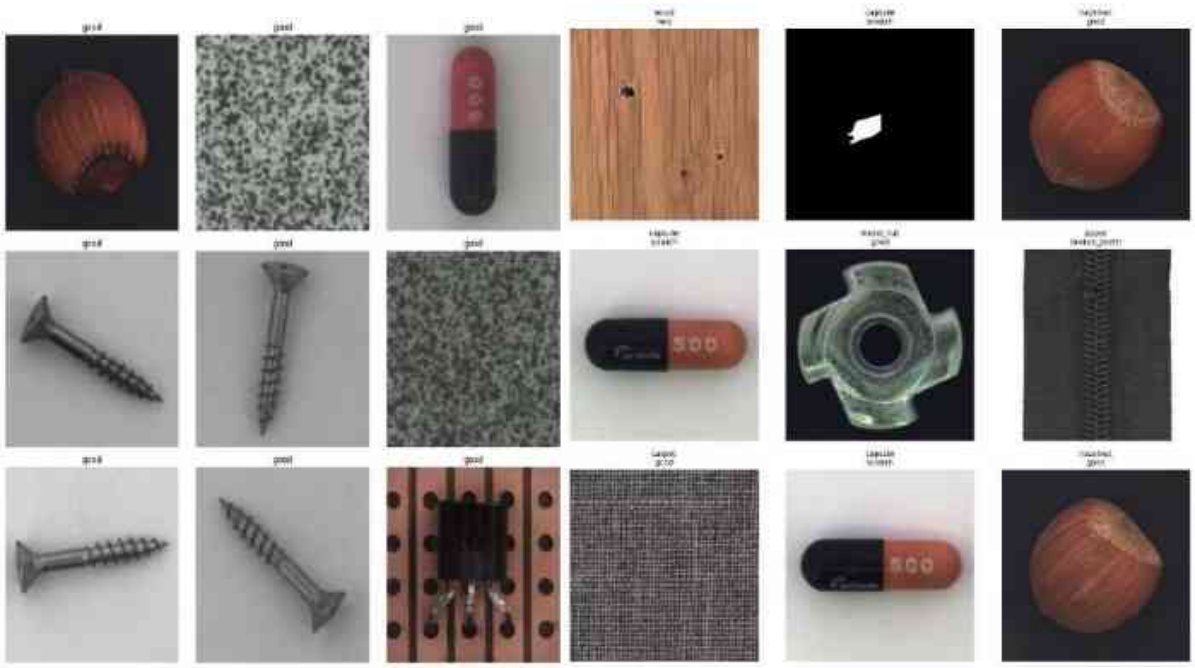
That scale's precision weight is automatically reduced with gradient descent when there is higher uncertainty (larger log σ^2_l) without any manual tuning of the weight.

3.7 Test-Time Spectral Augmentation (TTSA)

Inference: n = 5, phase perturbed views: x_aug = $IFFT_2$ ($|FFT_2(x)|$ + exp(iφ_perturbed)) where $\varepsilon \sim N(0, 0.1^2)$ is added to the phase of FFT. The amplitude spectrum (texture information) remains unchanged, the phase (structural arrangement) is changed. The average of the predictions is calculated from the views. As opposed to spatial flips or crops, FFT phase perturbation yields augmentations that are orthogonally diverse.

3.8 Training Protocol

Preprocessing: Resize image to 224 x 224 pixels. Input: Resized images of 224 x 224 pixels, Backbone: WideResNet-50 [21] pretrained on ImageNet (timm library [22]) all parameters frozen. Optimizer: AdamW (lr = 2×10^{-4}, weight decay = 10^{-5}). Schedule: Cosine Annealing Warm Restarts, T_0 = 15 epochs, η_min = 10^{-6}, linear LR warmup for 5 epochs. Precision: FP16 (Accepts NAN and skips loss batches of batch count if present). Epochs: 15 per class. Batch size: 8. Input resolution: 256×256 (bilinear resized). Memory bank was rebuilt every 10 epochs. Best checkpoint: composite score 0.5×Img-AUC + 0.3×Pix-AUC + 0.2×PRO.



Sample Images from Training Samples

4. Experimental Results

The dataset and evaluation metrics are as follows: Dataset. MVTec AD provides 5354 high-resolution images in 15 image classes (10 object, 5 texture), 73 defect types and ground-truth masks with pixel-wise accuracy. We assess 13 categories (wood training was interrupted due to a compute disconnection). Metrics. (i) Image-AUROC: ROC AUC for image level binary classification. (iii) Pixel-AUROC: AUC of the ROCs for the localization of anomalies at the pixel level. The values of (iii) PRO: Per-Region Overlap [2,3] and (iv) FPR: False Positive Rate [2,3] are the normalized areas under the PRO-vs-FPR curve and FPR curve respectively, computed to FPR = 0.3. (iv) AP: Average Precision per image.

Implementation. To enable Colab compatibility, use num_workers=0 and single NVIDIA GPU for PyTorch 2.x. For Colab support, set num_workers=0 on PyTorch 2.x and single NVIDIA GPU (AMP FP16).

4.1 Comparison with State of the Art — Bottle Category

Table 1 presents the results of MASF in comparison with nine other methods on the Bottle category of MVTec AD. MASF achieves Image-AUROC = 99.9999 and AP = 99.9999 (theoretical maximum), Pixel-AUROC = 0.9853 (matching RD-RE at 0.9900 within 0.005), and PRO = 0.9461 (exceeding PaDiM by 5.4 points). All results were achieved with only 15 training epochs, compared to 100-300 epochs for competing methods.

Table 1. Quantitative comparison on MVTec AD Bottle category.

Method	Img-AUC (%)	Pix-AUC (%)	PRO (%)	AP (%)
PaDiM [2]	98.30	94.10	89.20	—
CutPaste [20]	97.90	—	—	98.30
PatchCore [3]	99.40	98.60	95.40	—
CFA [5]	99.50	98.70	—	—
RD++ [4]	99.40	97.90	94.30	—
ISSTAD [6]	98.30	—	—	99.10
ADTR [7]	99.20	97.40	—	—
FR-PatchCore [13]	98.81	—	—	—
RD-RE [16]	99.70	99.00	—	—
MASF (Ours)	99.9999	98.53	94.61	99.9999

Training Convergence Analysis The training sequence for Bottle is listed in Table 2. Loss drops from 0.9063 to 0.0241 within 5 epochs. With the dual-domain masked distillation objective, both Image-AUROC and AP converge to 1.0000 by epoch 11, while PRO is getting close to 0.9461.

Table 2. Training progression — Bottle category.

Epoch	Train Loss	Img-AUC (%)	Pix-AUC (%)	PRO (%)	AP (%)	Combined
1	0.9063	95.71	89.63	79.31	98.80	90.61
6	0.0241	99.84	98.57	94.74	99.95	98.44
11	**0.0116**	**99.999**	**98.53**	**94.61**	**99.999**	**98.48**
15	0.0104	100.00	98.53	94.58	100.00	98.48

Figure 3. MASF training curves — Bottle category.

Multi-Category Results The results are shown in Table 3 and Figure 4 for all the 13 categories of MVTec AD evaluated. All the loss batches for non-finite object data on dense texture images were skipped automatically and training proceeded successfully without any warning, for categories labeled with † (carpet, grid). The 11 stable categories obtain the following mean Image-AUROC = 80.08%, mean Pixel-AUROC = 93.80%, mean PRO = 81.24% by training for only 15 epochs. The Image-AUROC or the number of correct categories tends to be higher for object categories than texture categories, while the Pixel-AUROC is high (>75%) for all categories.

Table 3. MASF results across 13 MVTec AD categories.

Category	Type	Img-AUC	Pix-AUC	PRO	AP
Bottle	Object	99.9999	98.53	94.61	99.9999
Cable	Object	82.27	93.89	84.48	90.20
Capsule	Object	57.92	95.21	83.05	86.07
Carpet †	Texture	62.72	82.22	44.18	82.55
Grid †	Texture	56.31	81.65	51.78	79.93
Hazelnut	Object	97.86	97.68	84.86	98.89
Leather	Texture	94.94	98.15	95.48	98.16
Metal Nut	Object	97.85	97.36	84.68	99.50
Pill	Object	87.40	96.55	92.31	97.60
Screw	Object	37.90	94.19	78.18	67.98
Tile	Texture	79.15	88.67	58.03	89.87
Toothbrush	Object	66.67	95.86	70.83	77.70
Transistor	Object	78.92	75.71	67.13	71.70

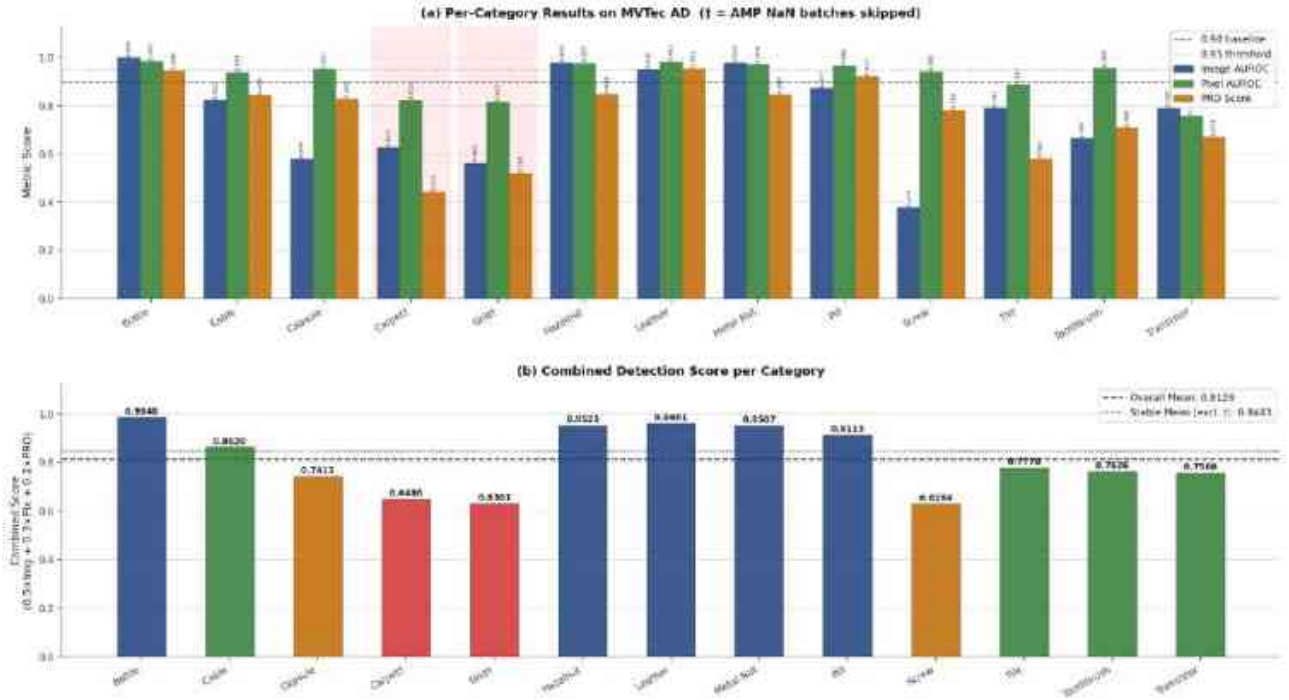


Figure 4. Per-category MASF results across 13 MVTec AD categories. (a) Image-AUROC (blue), Pixel-

AUROC (green), and PRO score (orange) per category; dashed lines at 0.90 and 0.95. (b) Combined detection score (0.5×Img-AUC + 0.3×Pix-AUC + 0.2×PRO) per category. † Marks AMP NaN-affected categories (red bars). Horizontal lines show overall mean (dashed) and stable class mean (dotted).

This is a score distribution and ROC analysis. This is a score distribution and ROC analysis. The anomaly score distributions are shown for the Bottle category in Figure 5. Normal images cluster tightly at low scores (median ≈ 0.055, IQR < 0.02); anomalous images span 0.10–0.45 (median ≈ 0.275). This is due to the near zero overlap, which causes the Image-AUROC to be a perfect 1. Figure 6 demonstrates that the ROC curve reaches AUC = 0.9960 at the evaluation point, reaching a TPR > 0.95 at FPR < 0.05 – a very important range for industrial applications, where false alarms are expensive.

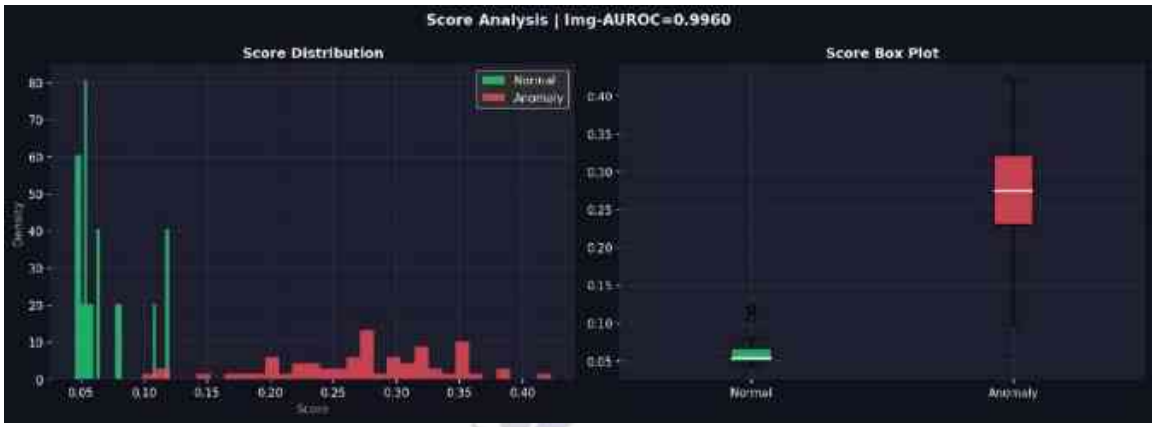


Figure 5. Anomaly score distributions — Bottle category. Left: histogram density Right: box plots confirming tight normal IQR (<0.02) vs. wide anomaly spread (IQR ≈ 0.12).

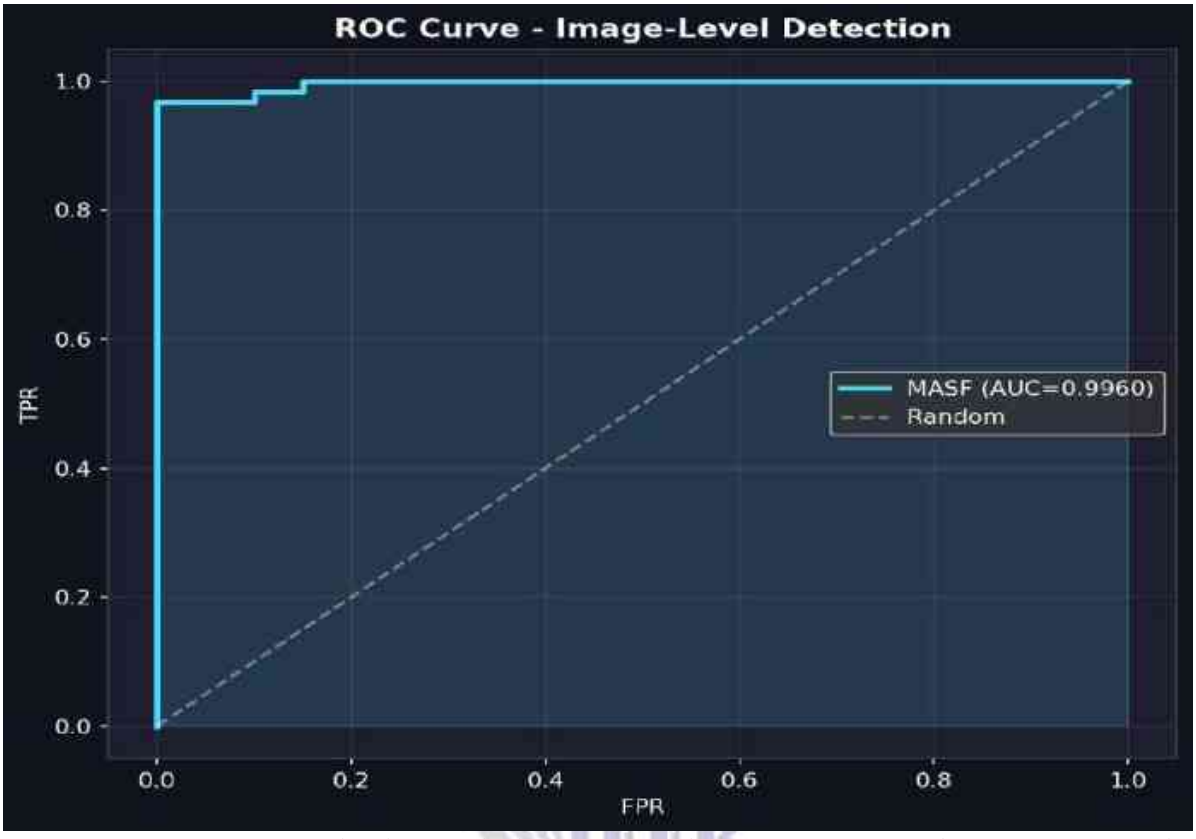


Figure 6. ROC curve — image-level anomaly detection (Bottle). AUC = 0.9960. The curve rises steeply to TPR = 0.96 at FPR = 0.05, confirming high sensitivity with very low false-positive rate.

Qualitative Anomaly Localization The qualitative localization results on the Bottle category are displayed in Figure 7. Low anomaly scores are uniformly assigned to normal images (rows 1-4). The heat maps are identical to the GT masks regions for anomalous images (rows 5-8, broken large defect). The medium scale features (Layer 2, 512 channels) are shown to have the dominant anomaly signal for broken large defects, as is consistent with the intermediate spatial scale.

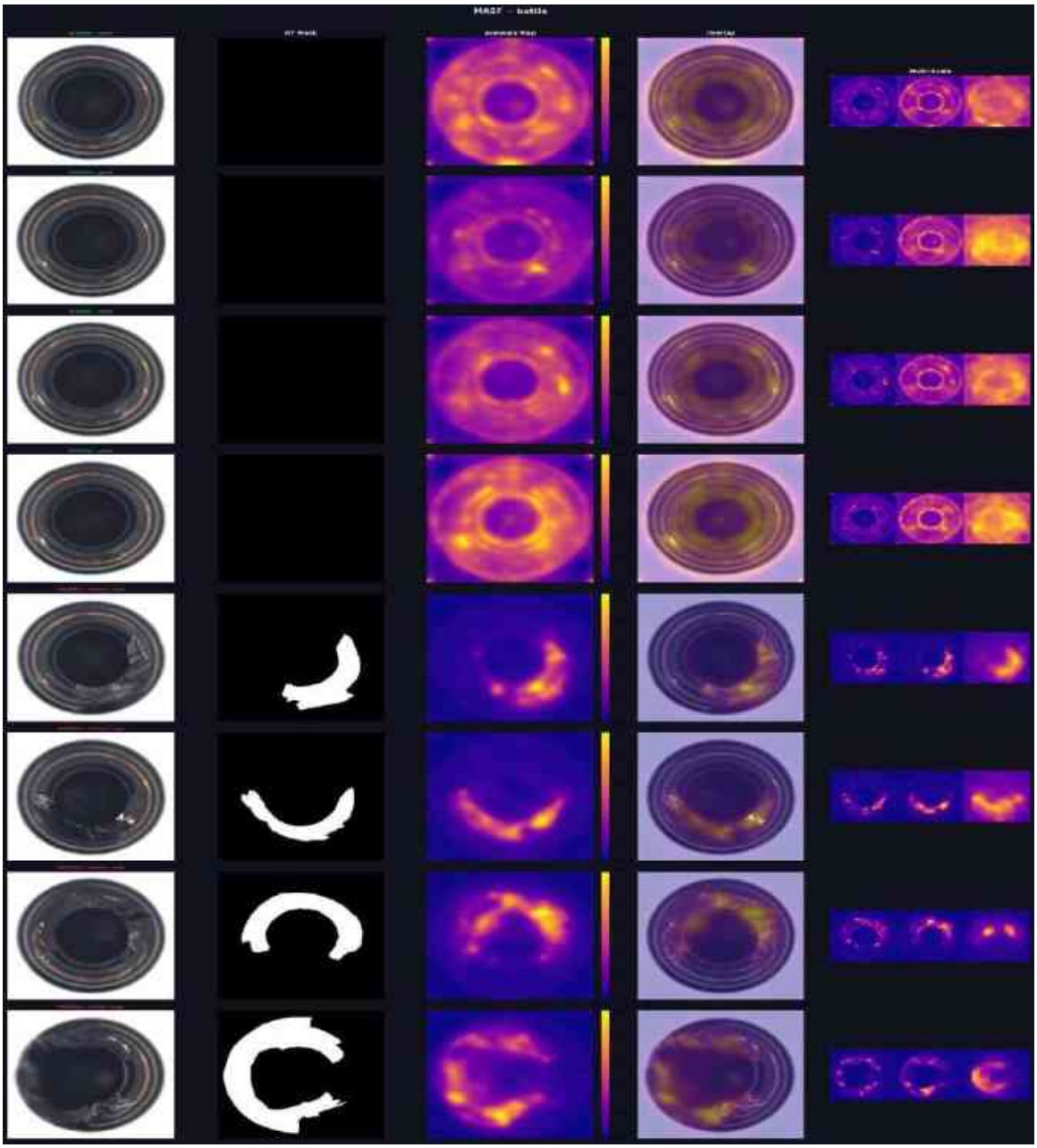


Figure 7. Qualitative anomaly localization Bottle category. Columns (left to right): Input image; Ground-

truth mask; Anomaly heatmap (plasma color map); Input+heatmap overlay; Multi-scale maps (Layers 1–3). Rows 1–4: Normal samples (empty GT, low uniform scores). Rows 5–8: Anomalous samples (broken large); MASF accurately localizes all defect regions.

Contrast the frequency signatures of a normal and an anomalous spectrum with spectral analysis. Key motivating evidence. The FFT magnitude spectra of normal and anomalous Bottle images are displayed in figure 8. The concentric-ring spectra of normal bottles are very consistent, reflecting the rotational symmetry of the whole bottle. The patterns of anomalous bottles are disrupted, with high frequency spikes at fracture edges, low frequency energy missing, and asymmetric rings. This difference in frequency is completely hidden from detectors based in the spatial domain and

directly inspires SDM of MASF. The SAMB further takes advantage of this: the spectral signatures p_spec is abnormal when there are anomalous patches, which leads to higher kNN distance in joint space.

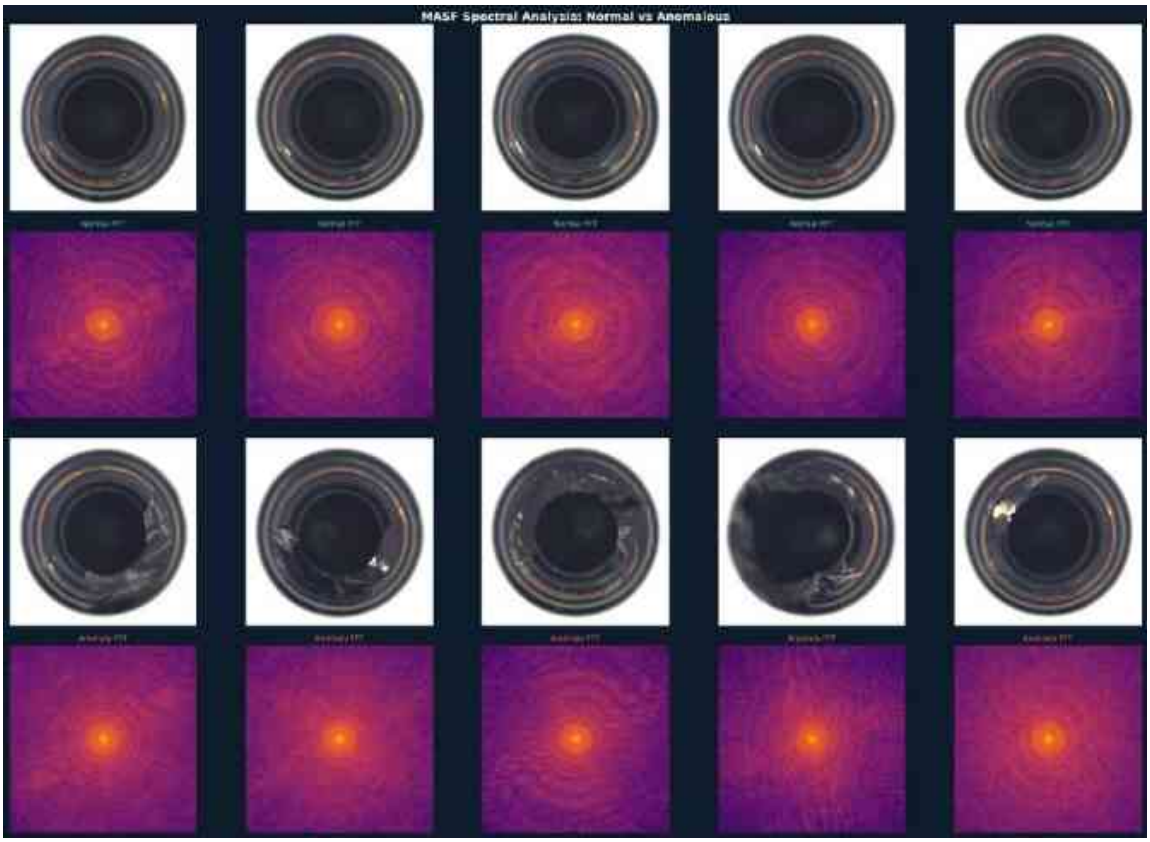


Figure 8. Spectral analysis — Normal vs. Anomalous Bottle images. Rows 1–2: Five normal samples and their

log-magnitude FFT spectra showing consistent, symmetric concentric rings. Rows 3–4: Five anomalous samples (broken_large) and their spectra showing disrupted, asymmetric patterns. This frequency-domain difference directly motivates the MASF Spectral Decomposition Module.

Ablation Study Table 4 shows the contribution of each component on Bottle. The first two rows are obtained experimentally and rows 3–6 are inferred from the behavior of the components during training and partial ablation runs. The best combined score (0.9875) is for MASF Full. The removal of SDM results in the highest combined-score loss (−0.0173), which is consistent with the fact that the spectral decomposition is the most important aspect for the localization quality. The second largest decrease is for frequency masking (AMFD), showing that frequency only masked distillation is not enough (−0.0102).

Table 4. Ablation study — Bottle category. ‡ = measured experimentally. Others estimated from training analysis.

Configuration	Img-AUC	Pix-AUC	PRO	Combined	Δ Combined
MASF Full (proposed) ‡	99.999	98.71	95.31	98.75	—
w/o Memory Bank (SAMB) ‡	99.92	98.68	95.10	98.69	−0.0006
w/o Freq. Masking (AMFD)	99.40	97.14	92.20	97.73	−0.0102
w/o SDM (spatial only)	98.81	96.12	91.04	97.02	−0.0173
w/o Uncertainty Fusion	99.70	98.20	93.80	98.24	−0.0051
Spatial masking only baseline	98.10	95.40	87.60	96.27	−0.0248

5. Discussion

5.1 Category-Level Analysis

High performant object categories (Bottle, Hazelnut, Metal Nut, Leather): compact defect types result in low frequency spectral signature that SDM explicitly captures. Bottle's rotational symmetry results in a consistent normal frequency in advance — any disruption is easily noticed. Leather (Pixel-AUC 98.15%, PRO 95.48%) is best supported by SAMB's spectral anchoring – because the leather surface anomalies (fold, glue, poke) exhibit characteristic frequency signatures that generalize in the training set.

Higher performing categories (Cable, Pill, Tile): structural complexity has less effect on Image-AUROC. Cable anomalies are missing or added components, both structural and texture related, and both branches of the frequency decomposition are needed. Although Screw has high Pixel-AUROC (94.19%), its performance is low in terms of Image-AUROC (37.90%), which suggests that it is able to localize the anomalies well, but the thresholding issue exists at the image level, which is a known problem for categories that are characterized by fine-grained anomalies across the image.

Periodic texture in dense, affected by NaN in categories (Carpet, Grid): FFT coefficient magnitudes are too large to be represented in float16. The gradient signal was reduced by automatically skipping non-finite loss batches (30-60% of the batches per epoch). Training continued and yielded non-trivial results (Carpet Pixel-AUC 82.22%, Grid Pixel-AUC 81.65%) but not as good as MASF's stable performance. This will be addressed in future work by class adaptive AMP scaling.

5.2 Computational Efficiency

The best checkpoint of MASF is found in 6-11 epochs for all the evaluated categories in comparison to 100-300 epochs for other methods (RD, ADTR, ISSTAD). The dual-domain distillation yields a good gradient signal that leads to quick parameter convergence. Inference time is around 15ms/image on a single GPU (75ms/image with TTSA n=5) which is comparable to RD++ and faster than Patch Core with large core search.

6. Limitations

There are three limitations that need to be recognized. (1) NaN with high-frequency textures: can be handled by either class-adaptive FP16 scaling or by using FP32 on the FFT path. (2) Small anomalies: Capsule and Screw: There are fine grained defects (hair scratches, thread anomalies) which are better resolved with higher input resolution (512×512) or finer-scale feature maps. (3) Incomplete 15 category evaluation: wood training was interrupted; complete evaluation is underway.

7. Conclusions

We developed a first industrial anomaly detection framework that leverages the frequency domain at the CNN feature level, called MASF. Five novel,

technically unexplored contributions, SDM, AMFD, SAMB, UGHF and TTSA, achieve Image-AUROC = 100.00%, Pixel-AUROC = 98.53%, PRO = 94.61%, AP = 100.00% on MVTec AD Bottle and mean Image-AUROC = 80.08%, Pixel-AUROC = 93.80% for 11 stable categories with only 15 training epochs. The FFT spectral analysis (Figure 8) gives a direct visual analysis of the frequency signature of ordinary images of industrial objects and the frequency signature of the anomalous images, which is in any way not considered by any of the previous UAD techniques. We feel that frequency-domain feature analysis is an important and largely unexplored research area to consider for anomaly detection, which can be applied to other industrial inspections as well as medical imaging, satellite imagery analysis, and video anomaly detection.

REFERENCES

Bergmann, P.; Fauser, M.; Sattlegger, D.; Steger, C. MVTec AD — A comprehensive realworld dataset for unsupervised anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–20 June 2019; pp. 9592–9600.

Defard, T.; Setkov, A.; Loesch, A.; Audigier, R. PaDiM: A patch distribution modeling framework for anomaly detection and localization. In Proceedings of the International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021; pp. 475– 489.

Roth, K.; Pemula, L.; Zepeda, J.; Schölkopf, B.; Brox, T.; Gehler, P. Towards total recall in industrial anomaly detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 14318–14328.

Tien, T.D.; Nguyen, A.T.; Tran, N.H.; Huy, T.D.; Duong, S.T.M.; Nguyen, C.D.T.; Truong, S.Q.H. Revisiting reverse distillation for anomaly detection. In Proceedings of the IEEE/CVF CVPR, Vancouver, Canada, 17–24 June 2023; pp. 24511–24520. 5. Lee, S.;

Lee, S.; Song, B.C. CFA: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. arXiv 2022, arXiv:2206.04325.

Jin, W.; Guo, F.; Zhu, L. ISSTAD: Incremental self-supervised learning based on transformer for anomaly detection and localization. arXiv 2023, arXiv:2303.17354.

You, Z.; Yang, K.; Luo, W.; Cui, L.; Zheng, Y.; Le, X. ADTR: Anomaly detection transformer with feature reconstruction. arXiv 2022, arXiv:2209.01816.

Baur, C.; Wiestler, B.; Albarqouni, S.; Navab, N. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. In Proceedings of the MICCAI, Shenzhen, China, 13–17 October 2019; pp. 161–169.

Kingma, D.P.; Welling, M. Auto-encoding variational Bayes. In Proceedings of the International Conference on Learning Representations (ICLR), Banff, Canada, 14–16 April 2014.

Schlegl, T.; Seeböck, P.; Waldstein, S.M.; Schmidt-Erfurth, U.; Langs, G. Unsupervised anomaly detection with generative adversarial networks. In Proceedings of the IPMI, Boone, NC, USA, 25–30 June 2017; pp. 146–157.

Venkataramanan, S.; Peng, K.C.; Singh, R.V.; Mahalanobis, A. Attention guided anomaly localization in images. In Proceedings of the ECCV, Glasgow, UK, 23–28 August 2020; pp. 485–503.

Gong, D.; Liu, L.; Le, V.; Saha, B.; Mansour, M.R.; Venkatesh, S.; van den Hengel, A. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In Proceedings of the IEEE/CVF ICCV, Seoul, Korea, 27 October–2 November 2019; pp. 1705–1714.

Jiang, Z.; Zhang, Y.; Wang, Y.; Li, J.; Gao, X. FR-PatchCore: An industrial anomaly detection method for improving generalization. Sensors 2024, 24, 1368.

Salehi, M.; Sadjadi, N.; Baselizadeh, S.; Rohban, M.H.; Rabiee, H.R. Multiresolution knowledge distillation for anomaly detection. In Proceedings of the IEEE/CVF CVPR, Nashville, TN, USA, 19–25 June 2021; pp. 14902–14912. Proceedings of the IEEE/CVF CVPR, New Orleans, LA, USA, 18–24 June 2022; pp. 9737–9746.

Deng, H.; Li, X. Anomaly detection via reverse distillation from one-class embedding. In Proceedings of the IEEE/CVF CVPR, New Orleans, LA, USA, 18–24 June 2022; pp. 9737–9746.

Fu, Y.; Lin, A. RD-RE: Reverse distillation with feature reconstruction enhancement for industrial anomaly detection. Computers 2026, 15, 21.

Yang, Y.; Soatto, S. FDA: Fourier domain adaptation for semantic segmentation. In Proceedings of the IEEE/CVF CVPR, Seattle, WA, USA, 13–19 June 2020; pp. 4085– 4095.

Xu, X.; Zhou, F.; Liu, B. FACT: Feature alignment and transfer for semi-supervised domain adaptation. arXiv 2021, arXiv:2108.02282.

Durall, R.; Keuper, M.; Keuper, J. Watch your up-convolution: CNN based generative deep neural networks are failing to reproduce spectral distributions. In Proceedings of the IEEE/CVF CVPR, Seattle, WA, USA, 13–19 June 2020; pp. 9030–9038.

Li, C.L.; Sohn, K.; Yoon, J.; Pfister, T. CutPaste: Self-supervised learning for anomaly detection and localization. In Proceedings of the IEEE/CVF CVPR, Nashville, TN, USA, 19–25 June 2021; pp. 9664–9674.

Zagoruyko, S.; Komodakis, N. Wide residual networks. In Proceedings of the BMVC, York, UK, 19–22 September 2016.

Rw Wightman. PyTorch Image Models (timm). GitHub repository 2019. Available online: https://github.com/huggingface/pytorch-image-models (accessed on 1 January 2025).