Demeaned Sparse: Efficient Anomaly Detection by Residual Estimate

Yifan Fang [* 1 2]   Yifei Fang [* 3]   Ruizhe Chen [* 1 4]   Haote Xu [1 2]   Xinghao Ding [1 2 4]   Yue Huang [1 2 4]

Abstract

Frequency-domain image anomaly detection methods can substantially enhance anomaly detection performance, however, they still lack an interpretable theoretical framework to guarantee the effectiveness of the detection process. We propose a novel test to detect anomalies in structural image via a Demeaned Fourier transform (DFT) under factor model framework, and we proof its effectiveness. We also briefly give the asymptotic theories of our test, the asymptotic theory explains why the test can detect anomalies at both the image and pixel levels within the theoretical lower bound. Based on our test, we derive a module called Demeaned Fourier Sparse (DFS) that effectively enhances detection performance in unsupervised anomaly detection tasks, which can construct masks in the Fourier domain and utilize a distribution-free sampling method similar to the bootstrap method. The experimental results indicate that this module can accurately and efficiently generate effective masks for reconstruction-based anomaly detection tasks, thereby enhancing the performance of anomaly detection methods and validating the effectiveness of the theoretical framework.

1. Introduction

Anomaly detection techniques, employed to capture anomalous structural changes within data that lead to deviations from nonanomalous distributions, are widely applied across various domains (Khan & Alkhathami, 2024; Jiang et al., 2023; Wu et al., 2024). The scarcity and diversity of anomalous data present significant challenges to collecting sufficient quantities for model training. Consequently, unsupervised anomaly detection methods, which utilize only nonanomalous samples as the training set, have emerged as a key research focus in this field (Zhang et al., 2024; Ristea et al., 2022; Reiss & Hoshen, 2023). Among these methods, reconstruction-based anomaly detection represents one of the mainstream approaches (Zavrtanik et al., 2022; 2021a; Madan et al., 2023). It detects anomalies by measuring differences between samples before and after reconstruction. The underlying assumption is that the model cannot accurately reconstruct anomalous structures not encountered during the training phase (Zavrtanik et al., 2021b).

Recently, Kascenas et al. (2023) proposed that anomalies can be viewed as noise added to the nonanomalous distribution, i.e., anomalous structures are considered an 'additive' phenomenon. Anomalies are detected by calculating the residuals between the reconstructed and original images. Furthermore, using residuals for structural change tests is very common in time series analysis, such as the Chow test (Toyoda, 1974), Ljung-Box test (Hassani & Yeganegi, 2020), and Breusch-Pagan test (Halunga et al., 2017). By the same way, anomalies in images can be considered a form of structural change, since the task can be regarded as a regression problem like the aforementioned tests. Fu et al. (2023) used a DFT-based method to detect structural changes in factor model, they gave a framework in factor level by capturing information from residuals which contain both the time and individual dimension. However, this test is not suitable for image anomaly detection tasks, as anomaly detection in images requires the application of a multi-dimensional DFT across two cross-sectional dimensions, which differs from the approach used in time series analysis.

In this paper, we follow the method from Fu et al. (2023) and incorporate a multi-DFT dual cross-sectional dimension of positional information as our primary method. we first demonstrate why anomaly detection can be effectively performed at the factor level using by capturing information from residuals. We also derive the test for such issue, as well as asymptotic properties of the statistics under null hypothesis and alternative hypothesis. The null hypothesis result let us compare with zero spectrum when there

---

[*]Equal contribution [1]Key Laboratory of Multimedia Trusted Perception and Efficient Computing, Ministry of Education of China, Xiamen University, Xiamen, China. [2]School of Informatics, Xiamen University, Xiamen, China. [3]School of Economics, Renmin University of China, Beijing, China. [4]Institute of Artificial Intelligence, Xiamen University, Xiamen, China. Correspondence to: Haote Xu <hotxu@stu.xmu.edu.cn>, Yue Huang <yhuang2010@xmu.edu.cn>.

is no anomaly, the global power and local power of our test show the effectiveness and the accuracy of our test, respectively. Our method reconsider the similarity between structure change and image anomaly detection. Following this motivation, we propose a general optimization module called the Demeaned Fourier Sparse (DFS) module, which is specifically designed to enhance reconstruction-based anomaly detection tasks.

Due to the advantage of residual-based test this module does not require consistent estimation of anomalous factors or prior knowledge of anomaly types to effectively detect anomalous samples. By applying our DFS module, we implemented the unsupervised anomaly detection task, the result shows our theoretical properties are robust, also the effectiveness of this module compared to the existing literature, which highlighting its strong capabilities in real-world applications.

The contributions of this paper are as follows.

- **Theoretical Contributions.** Inspired by the structural changes' detection of factor models, we propose an approach for detecting image anomalies by projecting residuals to a multi-fourier space. We derive the asymptotic properties of our test to ensure theoretical completeness, which demonstrates why anomaly can be detected at the factor level by using residuals. We also illustrate how this theory can be applied to reconstruction-based anomaly detection issues.

- **Practical Contributions.** We propose a factor model-based approach for reconstruction-based anomaly detection and have developed an effective module called Demeaned Fourier Sparse (DFS). Our method is simple and easy to implement. Our experimental results demonstrate the effectiveness of this approach, which is applicable to general reconstruction-based anomaly detection task.

The related work of factor models and unsupervised anomaly detection methods is presented in Appendix A. Section 2 introduces the hypothesis testing framework and statistical metrics. The asymptotic properties of the proposed method are detailed in Section 3. In Section 4, we design the DFS-AD method based on our testing framework. Finally, the experimental results presented in Section 5 demonstrate the accuracy of anomaly detection as well as the localization capabilities of DFS-AD.

2. Preliminaries

In this section, we propose a test-based approach in the factor model, which can detect anomalies at the factor level through residual analysis. The core concept involves applying the Discrete Fourier Transform (DFT) method to the factors and residuals from the factor model.

2.1. Hypothesis Testing

For an image $X \in \mathbb{R}^{H \times W}$, it can be interpreted as a sequence $\{X_{hw}, h = 1, 2, \ldots, H; w = 1, 2, \ldots, W\}$, consisting of observed values, where the (h, w)-th element represents the pixel located at the (h, w)-th position. We assume that X_{hw} is generated according to the following data generating process:

$$X_{hw} = \lambda_{hw}^{\top} F_w + \varepsilon_{hw}, \tag{1}$$

where F_w is a $R \times 1$ vector of unobserved common factors (To facilitate the expression of formulas and proofs, we assume that the direction of the common factor is consistent with W. In practical implementation, λ and F are mixed and difficult to distinguish, the factor direction can be arbitrary, this assumption does not affect our theory or proof), λ_{hw} is a $R \times 1$ vector of factor loadings that may vary across global or local features, R is the factor number, and ε_{hw} is the idiosyncratic error. The null hypothesis of no anomaly is given by:

$$\mathbb{H}_0 : \lambda_{hw} = \lambda_{h0} \text{ for all } h = 1, 2, ...H \text{ and all } w = 1, 2...W.$$

The alternative hypothesis is

$$\mathbb{H}_A : \lambda_{hw} \neq \lambda_{h0} \text{ for some } h \in \mathbb{H} \text{ and } w \in \mathbb{W},$$

where λ_{h0} represents a factor loading that is not anomalous. $\mathbb{W}$ and $\mathbb{H}$ denote the sets of indices (h, w) where $\lambda_{hw} \neq \lambda_{h0}$. The alternative hypothesis $\mathbb{H}_A$ encompasses various types of unknown loadings across different factors.

2.2. Test Statistic

Let $\varepsilon_{hw}^{\dagger} = (\lambda_{hw} - \lambda_{h0})^{\top} F + \varepsilon_{hw}$. Then (1) can be written as

$$X_{hw} = \lambda_{h0}^{\top} F_w + \varepsilon_{hw}^{\dagger}. \tag{2}$$

Let $\hat{F}_w$ and $\hat{\lambda}_{h0}$ denote the estimation of F_w and λ_{h0}, respectively. Then, the estimated residual is given by $\hat{\varepsilon}_{hw} = X_{hw} - \hat{\lambda}_{h0}^{\top} \hat{F}_w$.

Under the null hypothesis $\mathbb{H}_0$, $\varepsilon_{hw}^{\dagger} = \varepsilon_{hw}$ for all h and w, and (2) represents a normal pixel. Consequently, $\hat{F}_w$ and $\hat{\lambda}_{h0}$ are consistent estimators of the true common factors F_w and factor loadings λ_{h0}. Under the alternative hypothesis $\mathbb{H}_A$, we have: $\varepsilon_{hw}^{\dagger} = F_w^{\top}(\lambda_{hw} - \lambda_{h0}) + \varepsilon_{hw}$. In this expression, the first term contains anomalous information. Therefore, our goal is to estimate (2) and subsequently project the anomalous signals in the residuals into the spectral space.

Let $\hat{B} X_{hw} = \hat{\varepsilon}_{hw}$. It is straightforward to show that $\hat{B}$ is functioned as an annihilator matrix that project X_{hw} into the estimated residual $\hat{\varepsilon}_{hw}$. The following complex-valued empirical process can then be constructed:

$$\hat{C}(u,v) = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} \hat{F}_w \hat{\varepsilon}_{hw} \varphi(u,v)$$

$$= \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} \hat{F}_w \hat{B} X_{hw} \varphi(u,v)$$

$$= \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} \hat{F}_w F_w^\top \lambda_{hw} \hat{B} \varphi(u,v) \tag{3}$$

$$+ \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} \hat{F}_w \varepsilon_{hw} \hat{B} \varphi(u,v)$$

$$= C_1(u,v) + C_2(u,v),$$

where $\varphi(u,v)$ is a demeaned Fourier process, and $u,v \in \mathbb{R}$ are nuisance parameters such that $2\pi \left(\frac{u}{H} + \frac{v}{W} \right)$ represents a Fourier frequency. Here, combining the common factors and the residuals helps prevent the residuals from summing to zero.

$$\varphi(u,v) = e^{i2\pi \left(\frac{uh}{H} + \frac{vw}{W} \right)} - \frac{1}{HW} \sum_{m=1}^{H} \sum_{n=1}^{W} e^{i2\pi \left(\frac{um}{H} + \frac{vn}{W} \right)}.$$

Intuitively, after we project an image into the Fourier domain, we can interpret $C_1(u,v)$ as the "signal" corresponding to the anomalous factor loadings and its behavior is approximate to a zero spectral under $\mathbb{H}_0$, under $\mathbb{H}_A$ it will not. $C_2(u,v)$ as the "noise", contributing to a well-defined asymptotic distribution of the Discrete Fourier Transform (DFT) under $\mathbb{H}_0$. The anomalous characteristics of λ_{hw} are reflected through various patterns in the DFT within the frequency domain. Therefore, we can detect unknown types of anomalies by analyzing the deviation of the DFT from the zero spectrum across all frequencies.

In order to ensure the detection of unknown types of anomalies, it is necessary to estimate the order of magnitude that each $u, v \in \mathbb{R}$ deviates from $\hat{C}(u,v)$ and in order to facilitate the subsequent use of deep learning methods. The following test statistic is constructed:

$$\hat{D} = HW \iint_{\mathbb{R}} \|\hat{C}(u,v)\|^2 W(u,v) \, du \, dv. \tag{4}$$

The weight function $W(\cdot) : \mathbb{R} \to \mathbb{R}^+$ is a function designed to ensure the consistent integrability of $\hat{C}(u,v)$, and $\hat{D}$ is a measure used to evaluate the effectiveness of our test.

3. Asymptotic Theory

This section presents the asymptotic theory underlying our test statistic and explores some of its key properties. To address the practical considerations and requirements of anomaly detection tasks, we make the following assumptions:(i) Common factor: The high order moments of the factor is bounded, and its covariance matrix of factor

loading $\{\lambda_{hw}\}_{h=1,w=1}^{H,W}$ is positive definite, and the eigenvalues of the product of their covariance are distinct; (ii) The error term $\{\varepsilon_{hw}\}_{h=1,w=1}^{H,W}$ has weak dependence. (iii) ε_{hw} and F_w is strong mixing. (iv) the residual terms form a Martingale Difference Sequence. Such an assumption is appropriate for anomaly detection, where the objective is to identify irregular, random, and unpredictable anomalies. (v) The higher-order moments of the $W(\cdot)$ are finite, this guarantees the absolute integrability of the Fourier series. Under these assumptions, our goal is to maintain as mild conditions as possible. This approach ensures that the theoretical results remain valid under a broad range of conditions without imposing overly restrictive model specifications, thereby enhancing the applicability of this approach. Furthermore, it supports the analysis of anomaly detection tasks within the deep learning framework. Details of the proofs can be found in Appendix B.

3.1. Asymptotic Null Distribution

The asymptotic null distribution shows the properties of our test and its decision-making behavior under non-anomalous conditions. Under $\mathbb{H}_0$, when the factor loadings are non-anomalous, we assume that we choose the true number of factors.

Proposition 3.1. *Let $\delta = [-q, q]$ be a compact subset of $\mathbb{R}$ for any constant $q > 0$. Under the assumptions mentioned above and $\mathbb{H}_0$,*

$$\hat{C}(u,v) \Rightarrow H^{-1/2} W^{-1/2} G(u,v),$$

as $(H,W) \to \infty$, where $G(u,v)$ is a joint mean-zero joint complex-valued Gaussian process.

Proposition 3.1 demonstrates that under $\mathbb{H}_0$, the term $C_b(u,v)$, which represents the anomalous factor loading "noise" in $C(u,v)$ is dominant. For any compact set $\delta = [-q, q]$ with $q > 0$, $\sqrt{HW}\hat{C}(u,v)$ weakly converges to a joint mean-zero, joint complex-valued Gaussian process. The covariance kernel $\mathcal{K}$ is equivalent to the product of the asymptotic variance of $\frac{1}{\sqrt{HW}} \sum_{h=1}^{H} \sum_{w=1}^{W} \hat{F}_w \varepsilon_{hw} \hat{B} \varphi(u,v)$ and a pseudo-covariance induced by the Fourier transform.

Based on Proposition 3.1, the asymptotic distribution of the test statistic $\hat{D}$ under $\mathbb{H}_0$ is given by the following theorem.

Theorem 3.2. *When the conditions of Proposition 3.1 hold, under $\mathbb{H}_0$,*

$$\hat{D} \xrightarrow{d} \iint_{\mathbb{R}} \|G(u,v)\|^2 W(u,v) \, du \, dv.$$

as $(H,W) \to \infty$, where "$\xrightarrow{d}$" means convergence in distribution.

Theorem 3.2 presents the asymptotic null distribution of the test statistic $\hat{D}$. It is important to note that the integration

is carried out over $\mathbb{R}$. While Proposition 3.1 establishes weak convergence on any compact subset of $\mathbb{R}$, in practical applications, we use $\hat{D}$ to ensure that weak convergence holds as (u, v) tends to infinity.

3.2. Asymptotic Global Power

The asymptotic global power provides the properties and behaviors of our test under $\mathbb{H}_A$. It also offers an explanation of the asymptotic limit of the detection performance given by our test. As the number(size) of pixels(images) being analyzed increases, the detection task transitions from a pixel-level analysis to a more comprehensive image-level analysis.

The number of common factors selected under the alternative hypothesis $\mathbb{H}_A$ may not be the same as the true value, but our test still has power in most cases. At the image level, pixel-based anomaly detection tasks can be formulated as follows:

$$X = \mathbf{F}\mathbf{\Lambda}^\top + \varepsilon. \tag{5}$$

Here, $\mathbf{F}$ is the matrix of common factors, composed of F_w, and $\mathbf{\Lambda}$ is the factor loading matrix, composed of λ_{hw}. Next, We assume that the number of common factors to be estimated is K.

Proposition 3.3. *When $K < R$ for any compact set $\delta = [-q, q]$ with an arbitrarily large constant $q > 0$, as $(H, W) \to \infty$, $\sup_{u,v \in \delta} \left\| \hat{C}(u, v) - \mathcal{C}(u, v) \right\| = o_p(1)$, where $\mathcal{C}(u, v)$ denotes a nonzero spectrum.*

The non-zero spectrum provides the convergence of our test under $\mathbb{H}_A$, ensuring that the constructed complex-valued empirical process is significantly different to the zero spectrum. Specifically, this means that the estimator $\hat{C}(u, v)$ almost surely converges to the true value $C(u, v)$. Intuitively, the factor model is under-fitted when $K < R$. In such a case, the estimated residuals capture sufficient information about the unexplained common components. As a result, the DFT $\hat{C}(u, v)$ converges to a complex-valued process with non-zero mean value. Otherwise, when it is over-fitted with $K > R$, our test may has less power.

In practical applications, the goal is typically to reconstruct the complete image using as little information as possible. This implies estimating R with a smaller value of K, thereby promoting information sparsity. In practice, sparsity is often used to describe the relationship between the information utilized and the information reconstructed. Therefore, within the context of this analysis, we do not consider the case where $K > R$ in the asymptotic analysis.

Theorem 3.4. *When $K < R$, $P(\hat{D} > c_{HW}) \to 1$, as $(H, W) \to \infty$, for any $c_{HW} = o(HW)$.*

Theorem 3.4 presents the properties of the test statistic $\hat{D}$ under the alternative hypothesis $\mathbb{H}_A$. It indicates that when $K < R$, our test statistic $\hat{D}$ remains effective for image-level detection. Specifically, our power increases with the growth of HW, which enhances its ability to detect deviations from the distribution under $\mathbb{H}_0$, then increases the probability of rejecting the null hypothesis under $\mathbb{H}_A$. This theorem explains why, as the number of pixels (i.e., the image size) increases, our test is able to detect anomalies at the image level.

3.3. Asymptotic Local Power

The asymptotic local power provides an asymptotic lower bound for image anomaly detection. Even when the anomaly is very small(not under the lower bound), the residual-based test remains effective. We now consider the following setting for the anomaly factor loadings:

$$\mathbb{H}_A(\tau_{HW}) = \lambda_{hw} = \lambda_{h0} + \tau_{HW} g_{hw}. \tag{6}$$

Here, τ_{HW} is a scalar that depends on H and W which controls the strength of the anomalies, while g_{hw} represents the anomaly features for each pixel (h, w) it represents the components that deviate from the normal pixel. In general, λ_{h0} is not identifiable in the presence of anomalies, as the anomalies can obscure the true underlying factor loadings. Thus, we assume that $H^{-1}W^{-1}\sum_{h=1}^{H}\sum_{w=1}^{W} g_{hw} = 0$. It ensures that g_{hw} does not introduce a shift in the overall data, maintaining the balance of the data. In particular, it indicates that the impact of g_{hw} does not cause significant effects on λ_{h0}. This condition rules out the possibility of overestimating the number of common factors, as it ensures that the anomaly features g_{hw} cannot be eliminated through projection onto B. Intuitively, this indicates that the anomaly feature g_{hw} lies in the null space of λ_{h0}.

Proposition 3.5. *Let $\delta = [-q, q]$ be a compact subset of $\mathbb{R}$ for any constant $q > 0$. Then under assumptions given above and $\mathbb{H}_A(\tau_{HW})$ with $\tau_{HW} = H^{-1/2}W^{-1/2}$, as $(H, W) \to \infty$,*

$$\hat{C}(u, v) \Rightarrow H^{-1/2}W^{-1/2}(\phi(u, v) + G(u, v)),$$

where $\phi(u, v)$ denotes a pseudo-covariance between the Fourier series $e^{i2\pi\left(\frac{uh}{H} + \frac{vw}{W}\right)}$ and $g_{hw}B$. and $G(u, v)$ is as defined in Proposition 3.1.

$\tau_{HW} = H^{-1/2}W^{-1/2}$ provides an lower bound for the estimation with $K = R$. Proposition 3.5 implies that the asymptotic distribution of $\sqrt{HW}\hat{C}(u, v)$ under $\mathbb{H}_A(\tau_{HW})$ with $K = R$ is equivalent to a non-centralized complex-valued Gaussian process $\phi(u, v) + G(u, v)$. When there exists anomaly, which means $\phi(u, v) \neq 0$ for a non-negligible subset (u, v), $\hat{C}(u, v)$ can capture the anomalous behavior of the factor loadings λ_{hw}, provided that the subset is included in δ. To ensure the consistency of the DFT for all $u, v \in \mathbb{R}$, we have:

Theorem 3.6. *Under $\mathbb{H}_A(\tau_{HW})$ with $\tau_{HW} = H^{-1/2}W^{-1/2}$, $\phi(u,v)$ and $G(u,v)$ are as defined in Proposition 3.5 and Proposition 3.1*

$$\hat{D} \xrightarrow{d} \iint_{\mathbb{R}} \|\phi(u,v) + G(u,v)\|^2\, W(u,v)\, du\, dv,$$

as $(H,W) \to \infty$, where "$\xrightarrow{d}$" means convergence in distribution.

Theorem 3.6 provides the asymptotic distribution of $\hat{D}$ under the local alternative hypothesis $\mathbb{H}_A(\tau_{HW})$. This indicates that as long as the proportion of anomalies in $\lambda_{hw} - \lambda_{h0}$ is not too small, our test still has power. Intuitively, even if the anomalous characteristics of the factor loadings do not accumulate with the increasing sample size, the spatial information of a single sample still plays a crucial role in detecting anomalous changes. Since our test effectively leverages spatial information, it is capable of identifying anomalies at individual pixel locations. This indicates that our detection method is more sensitive to smaller anomalies. Furthermore, the theorem establishes the test's detection lower bound.

4. Methodology

Based on the aforementioned theory, we propose a hypothesis test along with its corresponding asymptotic theory. We argue that anomalies can be detected at the factor level through the analysis of residuals. Given the powerful fitting capabilities of deep learning networks, they are well-equipped to learn and approximate the factor structure of pixel-level data within the neural network. This allows for effective anomaly detection by leveraging the learned factor representations, facilitating the identification of deviations from the expected patterns in the data.

In this section, we revisit the theoretical foundations of anomaly detection tasks based on reconstruction methods. Reconstruction-based anomaly detection approaches effectively identify anomalies by accurately reconstructing nonanomalous samples during the inference phase, while anomalous samples fail to be reconstructed effectively. We propose the construction of a mask in the Fourier domain, which acts as an auxiliary mechanism for anomaly detection. The mask sparsifies the main information by reducing the number of estimated common factors and manipulates the image through element-wise multiplication. The mask construction follows our proposed progressive theory, and its selection is iteratively optimized using a joint loss function. Conceptually, this approach is similar to the bootstrap method commonly employed in non-parametric statistical techniques. DFS processes the input image X into $\hat{X}$, which is also an estimate of X. Subsequently, the output of reconstruction network X_{rec} is used to examine whether the λ
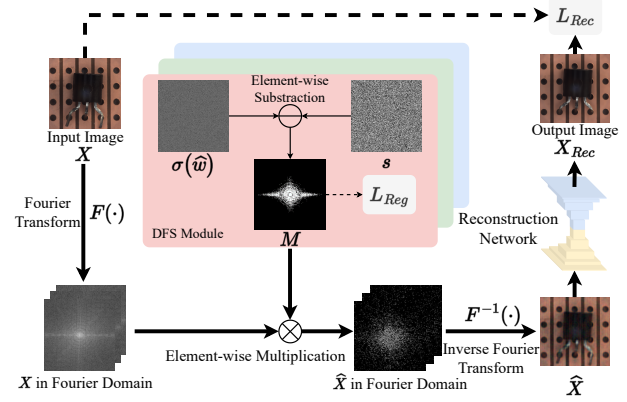

Figure 1. Overall framework of our method. The framework has two components: (1) DFS module is a specific implementation derived from the factor model theory, which adaptively learns Fourier space masks to construct the nonanomalous factors and residuals. (2) Reconstruction network aims at reconstructing normal regions. The predicted anomaly map $Pred = X - X_{Rec}$.

of each pixel in X belongs to λ_{h0}. Figure 1 illustrates the pipeline of our method:

Assume that each element M_{hw} in the mask matrix M follows a Bernoulli distribution, i.e., $M_{hw} \sim \text{Bernoulli}(p_{hw})$, where p_{hw} is obtained through iterative optimization within the module. Specifically, $M_{hw} = 1_A(x)$, where $1_A(x)$ is an indicator function, and A denotes the sample probability s_{hw}, with the condition $s_{hw} < p_{hw}$. This setup ensures that the mask elements are selected based on the probability p_{hw}, which is dynamically updated during the optimization process.

Sampling Method Firstly, we random sampling a tensor $\omega \in \mathbb{R}^{H \times W}$ from the [0,1] distribution and then map it to an optimizable variable $\hat{\omega}$ via the inverse function of the sigmoid function $\sigma(\cdot)$. Then sampling s_{hw} from $(0,1)$. In fact, there are no restrictions on the sampling distribution, referring to the Bootstrap method, the sampling distribution has no restrictions. Theoretically, any commonly used distribution on the interval $(0,1)$ can be applied.

Decentralization Thus, the probability map $p = \sigma(\hat{\omega})$ can be constructed using $\hat{\omega}$, where $\sigma(\cdot)$ is the sigmoid function. The mask elements $M_{hw} = 1_A(x)$ can then be approximated in the following optimizable form:

$$M = \sigma\left(\sigma(\hat{\omega}) - s\right), \quad s \sim (0,1). \tag{7}$$

Intuitively, as the number of iterations $T \to \infty$ increases, the sampling of s_{hw} during the iteration process will converge to the sample mean. This convergence ensures that the distribution of the sampled values stabilizes, aligning with the expected values. This concept is consistent with the Demeaned Fourier process introduced in Section 2.2, Our goal is to enable the model to optimize the residuals during

Table 1. Anomaly detection/localization results with Img-AUROC / Pix-AUROC / PRO (in %) metric on MvTec-AD. The best and second best performance are highlighted in bold and underline, respectively.

	Category	PatchCore	PyramidFlow	SSNF	DRAEM	IFgNet	TransFusion	Ours-Base	Ours
Texture	carpet	98.60 / 98.79 / 96.18	91.68 / 96.49 / 91.33	98.60 / 98.21 / 96.03	96.63 / 95.71 / 94.82	94.86 / 93.92 / 95.87	97.99 / 96.27 / 92.27	97.27 / 99.35 / 91.25	99.00 / 99.45 / 96.77
	Grid	98.12 / 98.31 / 95.60	86.17 / 96.11 / 92.78	99.52 / 96.10 / 95.04	99.99 / 99.34 / 97.89	100.00 / 99.19 / 92.30	100.00 / 99.58 / 98.51	99.67 / 99.08 / 93.12	100.00 / 98.90 / 94.91
	Leather	100.00 / 99.01 / 98.69	99.80 / 97.81 / 98.37	99.61 / 97.70 / 94.83	99.97 / 98.80 / 98.23	100.00 / 98.67 / 94.57	100.00 / 99.11 / 96.62	100.00 / 99.29 / 96.68	100.00 / 99.68 / 99.08
	Tile	98.91 / 94.12 / 87.80	99.39 / 95.58 / 88.80	99.90 / 94.43 / 87.82	99.92 / 99.39 / 97.10	99.93 / 99.22 / 95.33	100.00 / 99.65 / 98.80	99.35 / 97.77 / 90.79	100.00 / 98.69 / 95.13
	Wood	99.10 / 93.73 / 88.31	99.50 / 95.33 / 88.17	98.81 / 94.79 / 91.31	99.61 / 96.69 / 91.68	99.39 / 96.64 / 90.72	98.95 / 94.23 / 94.81	100.00 / 95.18 / 85.18	99.82 / 97.52 / 91.25
Object	Pill	96.29 / 97.58 / 93.60	81.49 / 93.58 / 76.52	98.09 / 98.62 / 94.63	97.50 / 97.48 / 91.41	96.54 / 96.33 / 91.38	98.34 / 98.03 / 97.92	91.84 / 91.63 / 79.71	99.59 / 98.79 / 97.64
	Transistor	100.00 / 92.94 / 81.03	92.88 / 96.71 / 75.43	94.42 / 96.11 / 79.12	92.38 / 86.22 / 72.19	94.58 / 77.71 / 62.48	99.54 / 77.04 / 83.22	96.12 / 93.45 / 72.49	99.42 / 95.32 / 84.98
	Cable	99.57 / 97.50 / 92.02	75.13 / 91.32 / 65.41	96.40 / 96.94 / 89.21	92.30 / 95.68 / 83.63	95.35 / 97.45 / 86.40	98.50 / 98.03 / 92.71	90.50 / 93.85 / 72.03	99.03 / 98.62 / 91.10
	Zipper	99.41 / 98.60 / 96.99	94.70 / 96.23 / 86.03	94.61 / 96.60 / 94.33	99.91 / 98.70 / 96.62	100.00 / 98.93 / 94.31	100.00 / 98.95 / 96.54	95.75 / 99.32 / 91.97	98.35 / 99.25 / 97.56
	Toothbrush	99.80 / 98.03 / 92.71	95.42 / 97.01 / 86.22	90.62 / 98.39 / 89.47	99.98 / 98.63 / 91.60	100.00 / 98.76 / 92.68	100.00 / 98.09 / 93.99	100.00 / 95.22 / 75.13	99.17 / 99.05 / 93.62
	Metal Nut	99.92 / 97.90 / 90.29	77.33 / 94.70 / 71.01	99.90 / 97.01 / 92.01	99.10 / 99.18 / 96.11	98.88 / 99.21 / 81.68	100.00 / 96.62 / 95.83	92.42 / 96.29 / 71.29	97.41 / 96.87 / 84.97
	Hazelnut	100.00 / 97.92 / 91.04	82.20 / 94.62 / 81.30	98.22 / 97.79 / 87.12	99.99 / 99.42 / 95.52	99.82 / 97.17 / 91.17	100.00 / 99.57 / 98.84	95.75 / 94.73 / 79.40	98.57 / 99.10 / 93.95
	Screw	98.01 / 99.42 / 97.90	53.44 / 86.10 / 56.11	88.80 / 98.03 / 93.59	95.70 / 98.61 / 96.49	92.60 / 97.24 / 93.05	98.63 / 99.17 / 96.79	69.62 / 97.63 / 90.44	96.02 / 99.23 / 96.69
	Capsule	97.73 / 98.70 / 96.41	94.21 / 98.40 / 93.92	94.21 / 97.79 / 94.38	97.02 / 92.88 / 90.50	94.69 / 98.88 / 94.51	99.40 / 97.96 / 94.74	97.25 / 97.78 / 87.67	94.10 / 96.25 / 88.18
	Bottle	100.00 / 98.03 / 95.80	85.70 / 95.79 / 79.13	100.00 / 95.81 / 86.03	98.59 / 99.21 / 96.39	96.51 / 95.96 / 90.66	100.00 / 99.00 / 97.81	100.00 / 98.30 / 84.39	99.92 / 98.05 / 92.75
	Average	98.96 / 97.37 / 92.96	87.27 / 95.05 / 82.04	96.78 / 96.95 / 90.99	97.91 / 97.06 / 92.68	97.54 / 96.49 / 89.81	99.42 / 96.75 / 95.29	95.04 / 96.59 / 84.10	98.69 / 98.32 / 93.24

the iteration process and use the mask to sparsify the main information.

Sparsification The obtained mask is applied to the test sample x in the Fourier domain, resulting in a masked image signal. Since the masking in the Fourier domain targets global information, it is equivalent to weakening the main information while increasing the residual, effectively introducing a sparsification operation. Such that $(K < R)$ ensures the validity of our test. Thus, a binary mask M for suppressing main information is obtained through decentralization, which adjusts the common factors in the Fourier space. The real and imaginary parts of the Fourier space after suppression are given by:

$$\hat{R}(x) = R(x) \odot M, \hat{I}(x) = I(x) \odot M, \qquad (8)$$

here, $\odot$ represents element-wise multiplication. The masked image is then transformed back into the image domain by the inverse Fourier transform. For the residual estimation, this highlights the deviations introduced by anomalies, allowing the detection method to more effectively distinguish between normal and anomalous patterns.

Optimization In this section, we use the iterative optimization process of the loss function to approximate the bootstrap resampling process. This substitution is conceptually similar as it transforms the simulation process into an iterative optimization problem. The training process is constrained by two main losses: the reconstruction loss L_{Rec} and the regularization loss L_{Reg}:

$$L = L_{Rec} + \alpha L_{Reg}, \qquad (9)$$

where $L_{Rec} = L_2 + L_{SSIM} + L_{MSGMS}$ includes the three terms, consisting of pixel-wise L_2 loss, structural similarity loss L_{SSIM} (Wang et al., 2004), and multiscale gradient magnitude similarity (MSGMS) loss (Xue et al., 2013)

L_{MSGMS}. The regularization loss $L_{Reg} = ||M||_1$, aiming at making the mask more sparse. The L_{Reg} is applied in the first half of the training process. In the second half of the training process, the mask is fixed, only the L_{Rec} is applied to continue training the reconstruction network.

5. Experiments

5.1. Experimental Settings

Datasets The experiments were performed on two widely used anomaly detection datasets: (1) MVTec-AD dataset (Bergmann et al., 2019a) serves as a benchmark for AD, which contains 10 object categories and 5 texture categories, totaling 5,354 colour images. (2) VisA dataset (Zou et al., 2022) is a large industrial anomaly detection dataset that contains 9,621 normal and 1,200 anomalous samples, covering 12 objects in 3 domains: Complex structure, Multiple instances and Single instance. Division of training and testing sets about the datasets can be found in Appendix C.

Implementation Details The framework were implemented in PyTorch (Paszke et al., 2019) and trained on two datasets with the same settings. Resolution of all input images are set to 256×256. The reconstruction network is a vanilla U-Net. Training epochs are set to 800, with batchsize of 2. Adam optimizer are used with an initial learning rate of 10^{-4}, the learning rate decays with a factor of 0.2 at the 640 and 720 epoch. The regularization coefficient α is set to 10^{-6}. The mask M is binarized at epoch 400 and fixed after binarization. The binarization strategy is to count N epochs before the mask is fixed, when the elements in M_{hw} exceed the threshold s_{hw}, where s_{hw} denotes the element in s which illustrated in Equation 7, its count is incremented by 1. When the 400th epoch ends, if the count of the M_{hw} element is greater than the preset sampling threshold multiplied by

Table 2. Anomaly detection/localization results with Img-AUROC / Pix-AUROC / PRO (in %) metric on VisA. The best and second best performance are highlighted in bold and underline, respectively.

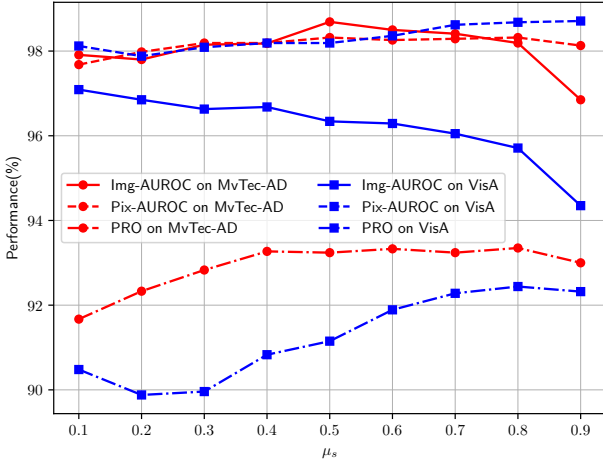|	Category	PatchCore	PyramidFlow	SSNF	DRAEM	IFgNet	TransFusion	Ours-Base	Ours
Complex	PCB1	<u>97.70</u> / **99.82** / <u>95.03</u>	85.19 / 98.02 / 72.88	91.12 / 99.40 / 92.69	78.42 / 85.30 / 25.64	95.29 / 98.70 / 78.53	**99.96** / 99.04 / **96.41**	97.08 / <u>99.45</u> / 86.89	97.46 / 99.42 / 93.04
	PCB2	95.31 / 97.80 / <u>91.62</u>	94.21 / 97.42 / 89.03	92.03 / 98.21 / 84.30	89.13 / 77.04 / 24.72	96.64 / 95.75 / 74.81	<u>99.06</u> / **98.98** / **93.27**	98.90 / 98.15 / 79.23	**99.22** / <u>98.70</u> / 87.33
	PCB3	97.14 / 98.43 / 93.21	81.33 / 98.20 / 91.39	88.24 / 98.15 / 87.42	89.60 / <u>98.83</u> / 91.50	94.89 / 97.82 / 91.51	<u>98.88</u> / 97.92 / **95.37**	96.30 / 98.71 / 87.66	**98.89** / **99.14** / <u>95.34</u>
	PCB4	<u>99.52</u> / 97.20 / 87.19	96.22 / 94.63 / 81.44	94.82 / 96.39 / 82.55	98.11 / 96.93 / 81.44	98.80 / **98.52** / 91.98	**99.75** / 97.05 / <u>92.71</u>	99.15 / 93.14 / 79.13	99.28 / 94.59 / 80.87
Multiple	Macaroni 1	94.41 / 99.12 / 96.72	82.80 / 98.72 / 93.89	95.10 / 99.53 / 96.12	88.52 / 99.70 / 97.41	<u>99.68</u> / <u>99.76</u> / <u>97.98</u>	**99.89** / <u>99.76</u> / 97.87	92.46 / 99.52 / 97.28	96.94 / **99.78** / **98.82**
	Macaroni 2	70.22 / 97.20 / 94.61	73.01 / 94.03 / 85.72	83.24 / 97.05 / 91.40	77.63 / 99.69 / 98.54	86.43 / **99.90** / 98.38	<u>98.43</u> / <u>99.80</u> / **98.99**	72.27 / 97.76 / 93.51	<u>90.08</u> / 99.74 / <u>98.81</u>
	Capsules	77.11 / 98.80 / 87.22	87.94 / 96.12 / 89.60	92.41 / 99.50 / 92.94	73.81 / 98.03 / 92.23	90.68 / 99.41 / 94.58	**98.70** / 99.77 / **99.37**	95.23 / <u>99.70</u> / 95.40	<u>95.92</u> / **99.87** / <u>98.19</u>
	Candle	**98.69** / **98.71** / **96.42**	84.03 / 87.22 / 75.54	90.50 / 98.01 / 91.30	91.04 / 95.24 / 87.52	92.93 / 97.17 / 87.94	<u>98.20</u> / <u>98.61</u> / <u>92.72</u>	95.89 / 93.78 / 75.95	88.44 / 95.84 / 89.20
Single	Cashew	<u>97.56</u> / 98.20 / <u>94.12</u>	93.40 / 95.93 / 89.31	87.54 / 98.20 / 90.82	90.20 / 78.33 / 45.24	90.72 / **99.19** / 91.18	92.82 / 96.54 / 94.66	**97.72** / <u>99.03</u> / 88.92	96.04 / 98.80 / 87.36
	Chewinggum	<u>99.22</u> / <u>98.63</u> / <u>91.03</u>	91.33 / 94.91 / 82.03	99.00 / **98.78** / 90.13	88.23 / 97.04 / 70.53	98.96 / 98.53 / 83.45	**99.98** / 98.36 / **91.44**	96.32 / 98.58 / 79.52	97.28 / 98.29 / 80.10
	Fryum	94.09 / 93.41 / <u>88.73</u>	91.84 / 94.02 / 87.78	97.70 / 92.42 / 78.61	86.20 / 91.63 / 86.04	92.68 / <u>96.61</u> / 84.05	**98.66** / **98.98** / 86.58	<u>98.36</u> / 95.20 / 81.94	97.84 / 94.71 / **88.93**
	Pipe fryum	<u>99.69</u> / 98.58 / <u>97.70</u>	76.93 / 96.68 / 88.12	95.13 / 98.22 / 92.70	94.72 / 93.11 / 90.32	97.78 / <u>99.28</u> / 93.74	**99.90** / 95.45 / <u>96.62</u>	98.84 / 98.54 / 88.23	98.68 / **99.41** / 95.82
	Average	93.39 / 97.99 / <u>92.80</u>	86.52 / 95.46 / 85.56	92.24 / 97.82 / 89.25	87.13 / 92.57 / 74.26	94.62 / <u>98.39</u> / 89.01	**98.69** / **98.44** / **94.67**	94.88 / 97.63 / 86.13	<u>96.34</u> / 98.19 / 91.15

Figure 2. Average performance of anomaly detection and localization under different μ_s.

the number of iterations during the statistical period, it is assigned to 1, otherwise 0, the preset sampling threshold is 0.5. All experiments were conducted on an NVIDIA RTX 3090 GPU. Details about the hyperparameter selection study can be found in Appendix D.

Compared Methods We compare our method with feature-based methods PatchCore (Roth et al., 2022), Pyramid-Flow (Lei et al., 2023), SSNF (Chiu & Lai, 2023), and Reconstruction-based methods DRAEM (Zavrtanik et al., 2021a), IFgNet (Chen et al., 2024), TransFusion (Fučka et al., 2024), in order to illustrate the effectiveness of DFS module, we added an extra setting for our method, that is, removing DFS from our framework (Noted as Ours-Base), only retaining the reconstruction network.

Evaluation Metrics We use the area under the receiver operating characteristic (ROC) curve at the testing stage as a substitute for the power of the test statistic $\hat{D}$ to evaluate the proposed method. This substitution is conceptually

straightforward as it allows us to translate the theoretical framework into an optimization problem suitable for training machine learning models. We evaluate our method using three metrics. Among them, the image-level area under the receiver operator curve (Img-AUROC) reports the performance of image-level anomaly detection. The pixel-level AUROC (Pix-AUROC) and per-region-overlap (PRO) report the performance and the accuracy of pixel-level anomaly localization, respectively. (PS:Note that statistical significance testing of the reported metrics is necessary, since the main contribution of this paper lies in the statistical theory. Moreover, because the method must be validated on practical detection problems, more comprehensive measures are required to assess overall performance. In particular, when verifying the asymptotic global theory (i.e., for the entire image), contributions from individual pixel p-values may be obscured (p-values are poorly adapted to imbalanced data, as they typically fail to account for anomaly sparsity). Therefore, we opt to use evaluation metrics commonly employed in the anomaly detection literature as our assessment criteria.)

5.2. Benchmarking

Table 1 presents a comprehensive summary of the anomaly detection and anomaly localization results on the MVTec-AD dataset. The average scores across all categories demonstrate that our method is competitive to most of the compared featured-based and reconstruction-based approaches. Table 2 provides the corresponding results on the VisA dataset, which are also competitive.

Both the MVTec-AD and VisA datasets feature complex structures and a diverse set of anomalies, highlighting the robustness of our approach and validating the theoretical foundations underpinning our method. To further assess the efficacy of the proposed DFS module, we report the results of the baseline model (Ours-Base) in Table 1 and
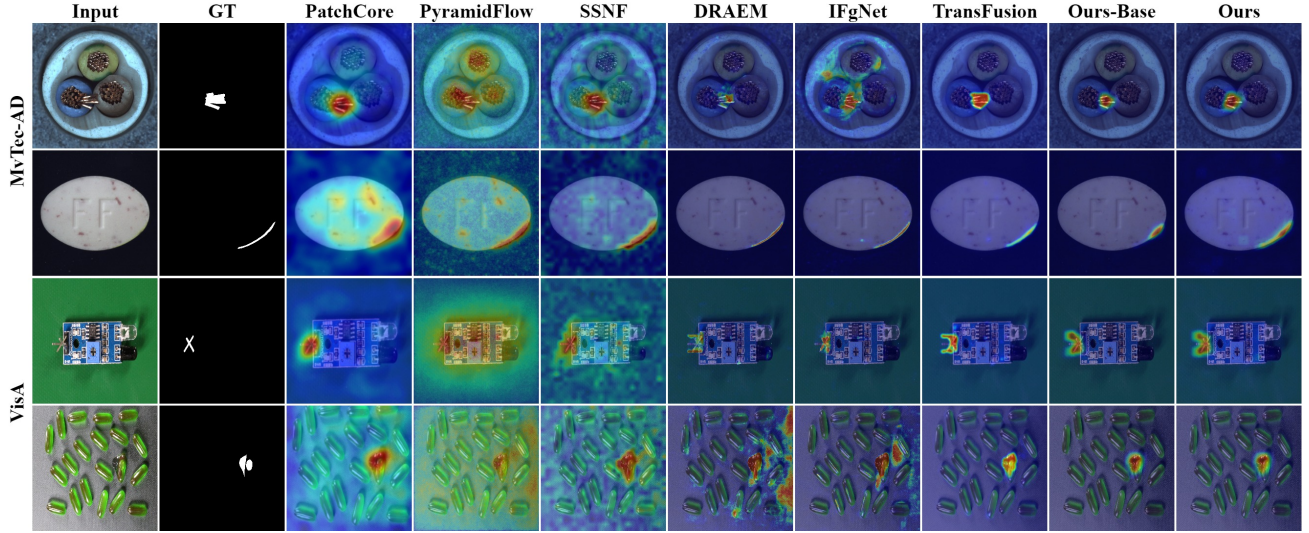
Figure 3. Qualitative comparison for anomaly localization on two datasets. From left to right: the input images, ground-truth, and the anomaly maps produced by all compared methods.

Table 2. The improvements achieved by our method over Ours-Base are 3.65%, 1.73%, and 9.14% for MVTec-AD, and 1.46%, 0.56%, and 5.02% for VisA, across the Img-AUROC, pix-AUROC, and PRO metrics, respectively. The PRO metric, being particularly sensitive to subtle anomalies, further substantiates the effectiveness of the asymptotic local power in the asymptotic theory we proposed.

To further validate our proposed asymptotic global power, we compared our method against baseline approaches on varying image sizes using the MVTec-AD dataset. The results in Table 3 demonstrate that, compared to the baselines, our method maintains high efficacy and exhibits greater stability as image size increases, thereby confirming the validity of the proposed asymptotic global theory.

In general, feature-based methods are typically fine-tuned using pretrained models, leading to relatively stable performance. In contrast, reconstruction-based methods often require data augmentation during the training process, and the external images introduced for augmentation tend to lack sufficient diversity, resulting in inconsistent performance across different datasets. Guided by the theory we have proposed, our method constructs a residual using the model itself to detect anomalies without the need for additional data, enabling precise anomaly detection. Furthermore, as shown in Table 4, our approach significantly reduces both model parameters and floating-point operations compared to other reconstruction-based methods.(PS:Note that the backbone of PatchCore is WideResNet50 (Zagoruyko, 2016), the coreset subsampling phase of PatchCore uses 1% feature vectors which are then indexed and stored in GPU memory to enable a fast search for nearest neighbors during inference. Therefore, the mechanism of PatchCore makes its Params and FLOPs varies with the amount of data, therefore
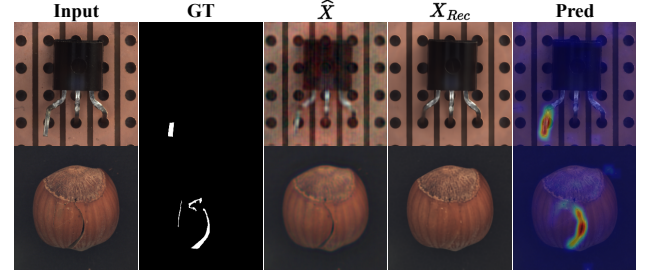

Figure 4. Visualization results of intermediate step in our method.

Table 4 only contains the computational cost of the layer used for computing features in Patchcore, the core subset of PatchCore and the kNN search is not reported.)

5.3. Hyperparameter Selection Study

Section 4 mentions that our resampling method is inspired by the Bootstrap method in nonparametric statistics, which does not assume a specific distribution for the signal. To evaluate the influence of the sampling function, we conduct experiments with different sampling functions, different means of mask sampling threshold s as well as the number of sampling epochs N to assess their impact on performance. The detailed results are provided in Appendix D.2. The results of sampling functions show that resampling using uniform and normal distributions yields nearly identical performance. The results of the mean mask sampling threshold s (μ_s) are shown in Figure 2. As μ_s changes from 0.1 to 0.9, the three metrics remain nearly identical trends across both datasets. This indicates that the DFS module is largely unaffected by the expected value of the sampling distribution. Additionally, results for varying the number of sampling epochs N indicate that performance remains stable as N in-

Table 3. Anomaly detection/localization results Img-AUROC / Pix-AUROC / PRO (in %) on MvTec-AD at different resolutions.

Resolution	256×256		512×512		1024×1024	
Category	Ours-Base	Ours	Ours-Base	Ours	Ours-Base	Ours
carpet	97.27 / 99.35 / 91.25	99.00 / 99.45 / 96.77	77.89 / 86.57 / 67.97	98.27 / 98.66 / 96.73	62.08 / 76.48 / 46.53	94.78 / 95.02 / 90.11
grid	99.67 / 99.08 / 93.12	100.00 / 98.90 / 94.91	97.33 / 97.49 / 79.79	99.67 / 99.29 / 96.86	96.74 / 91.22 / 61.34	100.00 / 99.40 / 97.69
leather	100.00 / 99.29 / 96.68	100.00 / 99.68 / 99.08	100.00 / 99.45 / 93.82	100.00 / 99.68 / 99.35	98.06 / 91.60 / 87.56	99.18 / 97.31 / 97.16
tile	99.35 / 97.77 / 90.79	100.00 / 98.69 / 95.13	96.83 / 96.67 / 86.95	100.00 / 98.58 / 96.16	93.65 / 88.87 / 73.62	98.70 / 96.57 / 92.84
wood	100.00 / 95.18 / 85.18	99.82 / 97.52 / 91.25	63.86 / 84.21 / 66.59	99.91 / 93.93 / 88.99	98.16 / 85.26 / 69.69	97.54 / 90.31 / 84.33
pill	91.84 / 91.63 / 79.71	99.59 / 98.79 / 97.64	82.38 / 74.72 / 58.93	99.78 / 98.55 / 96.99	77.93 / 58.43 / 40.51	96.15 / 94.91 / 79.84
transistor	96.12 / 93.45 / 72.49	99.42 / 95.32 / 84.98	87.46 / 83.69 / 46.11	97.96 / 87.09 / 76.14	60.38 / 62.66 / 29.21	95.38 / 75.71 / 74.58
cable	90.50 / 93.85 / 72.03	99.03 / 98.62 / 91.10	71.44 / 85.75 / 50.73	97.34 / 95.42 / 84.47	52.44 / 80.73 / 46.61	94.04 / 87.00 / 76.11
zipper	95.75 / 99.32 / 91.97	98.35 / 99.25 / 97.56	97.14 / 99.09 / 81.13	95.46 / 98.58 / 95.99	97.30 / 92.80 / 87.50	94.38 / 97.25 / 93.71
toothbrush	100.00 / 95.22 / 75.13	99.17 / 99.05 / 93.62	92.78 / 89.78 / 60.22	100.00 / 98.70 / 94.06	81.94 / 82.32 / 43.05	98.89 / 97.11 / 86.00
metal_nut	92.42 / 96.29 / 71.29	97.41 / 96.87 / 84.97	90.47 / 95.62 / 63.76	97.07 / 91.86 / 82.40	88.51 / 92.53 / 52.80	98.68 / 85.01 / 79.10
hazelnut	95.75 / 94.73 / 79.40	98.57 / 99.10 / 93.95	80.00 / 69.99 / 53.06	99.07 / 99.12 / 96.39	68.25 / 60.73 / 38.33	98.79 / 97.91 / 95.45
screw	69.62 / 97.63 / 90.44	96.02 / 99.20 / 96.69	51.98 / 96.26 / 87.43	96.74 / 97.82 / 93.77	63.05 / 95.91 / 85.01	96.18 / 96.32 / 89.82
capsule	97.25 / 97.78 / 87.67	94.10 / 96.25 / 88.18	87.16 / 92.80 / 74.78	97.85 / 88.76 / 88.71	87.91 / 91.44 / 77.31	96.22 / 67.43 / 79.97
bottle	100.00 / 98.30 / 84.29	99.92 / 98.05 / 92.75	100.00 / 97.48 / 67.35	98.73 / 95.98 / 88.37	98.41 / 91.53 / 42.78	98.89 / 95.06 / 85.56
average	95.04 / 96.59 / 84.10	98.69 / 98.32 / 93.24	85.11 / 89.97 / 69.24	98.52 / 96.13 / 91.69	81.65 / 82.83 / 58.79	97.19 / 91.49 / 86.82

creases from 5 to 35. To optimize computational efficiency, we select $\mu_s = 0.5$ and $N = 5$ for our experiments.

5.4. Qualitative Results

Some qualitative comparison results are presented in Figure 3. The proposed DFS-AD demonstrates visually accurate localization performance. Additionally, the intermediate steps in our method are visualized in Figure 4. The anomalous image X is initially masked in the Fourier domain using DFS module. The resulting image $\hat{X}$ is treated as the nonanomalous factor and subsequently reconstructed by the reconstruction network to generate the corresponding nonanomalous sample X_{Rec}. This approach effectively removes the abnormal residual, enabling precise anomaly detection and localization. Further qualitative results are provided in Appendix E.

Table 4. Computational cost evaluation of comparison methods. Number of model parameters (Params) and floating-point operations (FLOPs) are reported.

	PatchCore	PyramidFlow	SSNF	DRAEM	IFgNet	TransFusion	Ours
Params (M)	24.86	22.64	294.67	97.42	69.58	54.53	28.37
FLOPs (G)	9.24	479.20	243.77	198.66	135.23	148.38	38.20

6. Conclusion

In this work, we propose a novel method for detecting image anomalies using a demeaned Fourier transform (DFT)

and provide a theoretical demonstration of its effectiveness. The asymptotic theory elucidates why this test can detect anomalies at both the image and pixel levels within the theoretical lower bound. Then we discuss the relationship between the test and practical problems, we propose DFS as an optimized reconstruction anomaly detection module and develop the DFS-AD framework. Furthermore, we introduce a Bootstrap-like resampling method to optimize our framework. We quantitatively and qualitatively demonstrate the effectiveness of our method in reconstruction-based anomaly detection tasks.

Acknowledgements

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none which we feel must be specifically highlighted here.

References

Abdulaal, A., Liu, Z., and Lancewicki, T. Practical approach to asynchronous multivariate time series anomaly detection and localization. In *Proceedings of the 27th ACM SIGKDD conference on knowledge discovery & data mining*, pp. 2485–2494, 2021.

Bai, J. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003.

Bai, J. and Perron, P. Estimating and testing linear models with multiple structural changes. *Econometrica*, pp. 47–78, 1998.

Bergmann, P., Fauser, M., Sattlegger, D., and Steger, C. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 9592–9600, 2019a.

Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., and Steger, C. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*. SCITEPRESS-Science and Technology Publications, 2019b.

Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., and Steger, C. The mvtec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 129 (4):1038–1059, 2021.

Breitung, J. and Eickmeier, S. Testing for structural breaks in dynamic factor models. *Journal of Econometrics*, 163 (1):71–84, 2011.

Chen, X., Xu, H., Deng, C., Tu, X., Ding, X., and Huang, Y. Implicit foreground-guided network for anomaly detection and localization. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 2970–2974. IEEE, 2024.

Cheng, X., Liao, Z., and Schorfheide, F. Shrinkage estimation of high-dimensional factor models with structural instabilities. *The Review of Economic Studies*, 83(4): 1511–1543, 2016.

Chiu, L.-L. and Lai, S.-H. Self-supervised normalizing flows for image anomaly detection and localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshop*, pp. 2927–2936, 2023.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. The generalized dynamic-factor model: Identification and estimation. *Review of Economics and statistics*, 82(4):540–554, 2000.

Fu, Z., Hong, Y., and Wang, X. Testing for structural changes in large dimensional factor models via discrete fourier transform. *Journal of Econometrics*, 233(1):302–331, 2023.

Fučka, M., Zavrtanik, V., and Skočaj, D. Transfusion–a transparency-based diffusion model for anomaly detection. In *European conference on computer vision*, pp. 91–108. Springer, 2024.

Halunga, A. G., Orme, C. D., and Yamagata, T. A heteroskedasticity robust breusch–pagan test for contemporaneous correlation in dynamic panel data models. *Journal of econometrics*, 198(2):209–230, 2017.

Hassani, H. and Yeganegi, M. R. Selecting optimal lag order in ljung–box test. *Physica A: Statistical Mechanics and its Applications*, 541:123700, 2020.

Jiang, M., Hou, C., Zheng, A., Han, S., Huang, H., Wen, Q., Hu, X., and Zhao, Y. Adgym: Design choices for deep anomaly detection. *Advances in Neural Information Processing Systems*, 36:70179–70207, 2023.

Kascenas, A., Sanchez, P., Schrempf, P., Wang, C., Clackett, W., Mikhael, S. S., Voisey, J. P., Goatman, K., Weir, A., Pugeault, N., et al. The role of noise in denoising models for anomaly detection in medical images. *Medical Image Analysis*, 90:102963, 2023.

Khan, M. M. and Alkhathami, M. Anomaly detection in iot-based healthcare: machine learning for enhanced security. *Scientific Reports*, 14(1):5872, 2024.

Lei, J., Hu, X., Wang, Y., and Liu, D. Pyramidflow: High-resolution defect contrastive localization using pyramid normalizing flow. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14143–14152, 2023.

Li, C., Guo, C.-L., Liang, Z., Zhou, S., Feng, R., Loy, C. C., et al. Embedding fourier for ultra-high-definition low-light image enhancement. In *The Eleventh International Conference on Learning Representations*, 2023.

Lin, J., Gan, C., and Han, S. Tsm: Temporal shift module for efficient video understanding. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 7083–7093, 2019.

Madan, N., Ristea, N.-C., Ionescu, R. T., Nasrollahi, K., Khan, F. S., Moeslund, T. B., and Shah, M. Self-supervised masked convolutional transformer block for

anomaly detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

Newey, W. K. Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pp. 1161–1167, 1991.

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Perera, P., Nallapati, R., and Xiang, B. Ocgan: One-class novelty detection using gans with constrained latent representations. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2898–2906, 2019.

Rao, K. R., Kim, D. N., and Hwang, J. J. *Fast Fourier transform-algorithms and applications*. Springer Science & Business Media, 2011.

Reiss, T. and Hoshen, Y. Mean-shifted contrastive loss for anomaly detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 2155–2162, 2023.

Ristea, N.-C., Madan, N., Ionescu, R. T., Nasrollahi, K., Khan, F. S., Moeslund, T. B., and Shah, M. Self-supervised predictive convolutional attentive block for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 13576–13586, 2022.

Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., and Gehler, P. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14318–14328, 2022.

Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International conference on information processing in medical imaging*, pp. 146–157. Springer, 2017.

Stock, J. H. and Watson, M. Forecasting in dynamic factor models subject to structural instability. *The Methodology and Practice of Econometrics. A Festschrift in Honour of David F. Hendry*, 173:205, 2009.

Toyoda, T. Use of the chow test under heteroscedasticity. *Econometrica: Journal of the Econometric Society*, pp. 601–608, 1974.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Wu, P., Liu, J., He, X., Peng, Y., Wang, P., and Zhang, Y. Toward video anomaly retrieval from video anomaly detection: New benchmarks and model. *IEEE Transactions on Image Processing*, 33:2213–2225, 2024.

Xue, W., Zhang, L., Mou, X., and Bovik, A. C. Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE transactions on image processing*, 23(2):684–695, 2013.

Ye, F., Huang, C., Cao, J., Li, M., Zhang, Y., and Lu, C. Attribute restoration framework for anomaly detection. *IEEE Transactions on Multimedia*, 24:116–127, 2020.

Zagoruyko, S. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

Zavrtanik, V., Kristan, M., and Skočaj, D. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 8330–8339, 2021a.

Zavrtanik, V., Kristan, M., and Skočaj, D. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2021b.

Zavrtanik, V., Kristan, M., and Skočaj, D. Dsr–a dual subspace re-projection network for surface anomaly detection. In *European conference on computer vision*, pp. 539–554. Springer, 2022.

Zhang, K., Wen, Q., Zhang, C., Cai, R., Jin, M., Liu, Y., Zhang, J. Y., Liang, Y., Pang, G., Song, D., et al. Self-supervised learning for time series analysis: Taxonomy, progress, and prospects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

Zou, Y., Jeong, J., Pemula, L., Zhang, D., and Dabeer, O. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European Conference on Computer Vision*, pp. 392–408. Springer, 2022.

A. Related Work

Unsupervised Anomaly Detection Unsupervised Anomaly Detection (AD) can be categorized into reconstruction-based AD methods, flow-based AD methods, and feature-based AD methods. Reconstruction-based AD methods identify anomalies by comparing the differences between reconstructed images and the original images. To enhance the reconstruction accuracy of the network, Bergmann et al. (2019b) utilized the L_2 norm and the Structure Similarity Index Measure (SSIM) as the loss function. Furthermore, AnoGAN (Schlegl et al., 2017) constructs a latent space representing the distribution of normal images and employs an optimization-based approach to search for the closest embedding of the input image within this space, then the embedding is decoded to obtain the reconstructed image. In addition, DRAEM (Zavrtanik et al., 2021a) enhances the network's reconstruction capability through the use of synthetic anomalies. Moreover, this method trains a discriminator sub-network to detect anomalies. Considering the potential presence of background noise in real-world scenarios, IFgNet (Chen et al., 2024) adopts a multi-task framework that enables the network to simultaneously detect anomalies and extract foreground objects, which can eliminate interference effectively. Recently, to address the issues of overgeneralization and loss of detail commonly seen in reconstruction models, **?** proposed a method based on a transparency-based diffusion process that improves the detection accuracy significantly. Flow-based methods utilize pre-trained models to extract features from nonanomalous samples and model their probability distributions, attributing to the detection of out-of-distribution anomalies. A notable example is PyramidFlow (Lei et al., 2023), which employs pyramid-like normalizing flows and volume normalization to facilitate high-resolution anomaly localization. Similarly, Chiu & Lai (2023) introduced a self-supervised normalizing flow-based model that integrates synthetic anomalies into the flow-based framework. This approach conditionally optimizes the model by maximizing the likelihood of nonanomalous features while minimizing the likelihood of synthetic anomaly features, allowing the model to capture the distribution of nonanomalous features more accurately. Feature-based anomaly detection methods detect anomalies by comparing the differences in features between normal and abnormal images. PatchCore (Roth et al., 2022) extracts a core subset of patch-level features from training images using a pre-trained network. Anomaly scores are calculated by measuring the distance between the features of the test images and the core subset. OCGAN (Perera et al., 2019) proposes to add Gaussian noise to the input, while ARNet (Ye et al., 2020) uses some transformations to erase some important attributes in the images. However, they do not provide interpretability or theoretical underpinning for the method.

Factor Model Factor models are commonly used for analyzing structured data (Bai & Perron, 1998), such as in finance, economics (Forni et al., 2000), and weather forecasting. Most studies assume that factor loadings are time-invariant (Bai, 2003), capturing the relationship between observed data and unobserved common factors. In recent years, research in such field has increasingly focused on testing and estimating factor models with structural changes (Stock & Watson, 2009; Breitung & Eickmeier, 2011; Cheng et al., 2016). However, such tests and estimations often specify on a certain type of structural change, like abrupt structural changes or smooth structural changes. Fu et al. (2023) proposed tests for structural changes in time-varying factor loadings scenarios, considering both abrupt and smooth structural changes. In the context of common image anomaly detection tasks, the data can be considered structured (Bergmann et al., 2021), as these tasks usually assume that anomalies occur according to the practical scenario. Compared to economic data, this is reflected on dimensionality, structural changes in panel data can be mapped to anomalies in images. For example, a structural change corresponds to anomaly in certain pixel positions within the image, while in panel data it reflects the break in both cross-sectional level and time-series level.

Discrete Fourier Transform The Discrete Fourier Transform (DFT) is commonly used in image processing, with the goal of transforming image information into the spectral domain for further analysis (Rao et al., 2011). The coefficients in the frequency domain contain global information about the image (Lin et al., 2019). Signal decomposition methods can be employed to decouple the image's features in the frequency domain. Li et al. (2023) proposed an embedded Fourier enhancement method to improve low-light images, aiming to enhance illumination information at the feature level using Fourier methods. Abdulaal et al. (2021) introduced an asynchronous multivariate time series anomaly detection and localization method for asynchronous multivariate time series data. Based on the theory of factor models, we reconsider the theoretical construction of reconstruction-based anomaly detection by projecting the image features into the Fourier domain, representing the image information as a combination of signal and noise. In this way, we propose a simple and effective mask-based anomaly detection method.

B. Proofs

B.1. Proof of Proposition 3.1

Let $\Phi(u)$ be an $W \times W$ diagonal matrix, where the m-th diagonal element in $\Phi(u)$ is $\varphi(u)$, with $\varphi_{(m)}(u) = e^{i2\pi\left(\frac{um}{H}\right)} - \frac{1}{H}\sum_{h=1}^{H} e^{i2\pi\left(\frac{uh}{H}\right)}$ representing a demeaned Fourier process. Let $\Phi(v)$ be an $H \times H$ diagonal matrix, where the n-th diagonal element in $\Phi(v)$ is $\varphi(v)$, with $\varphi_{(n)}(v) = e^{i2\pi\left(\frac{vn}{W}\right)} - \frac{1}{W}\sum_{w=1}^{W} e^{i2\pi\left(\frac{vw}{W}\right)}$ representing a demeaned Fourier process. α is an $H \times 1$ vector with each entry being $\frac{1}{H}$. The matrix Λ_0 contains the set of non-anomalous factor loadings. B_{HW} is a matrix containing the K largest eigenvalues of the matrix $H^{-1}W^{-1}XX^{\top}$, and $A = H^{-1}W^{-1}(\Lambda_0^{\top}\Lambda_0)(F^{\top}\hat{F})B_{HW}^{-1}$ is a rotation matrix.

Based on the definition of $\hat{C}(u, v)$, under the null hypothesis $\mathbb{H}_0$,

$$\hat{C}(u,v) = \frac{\hat{F}^{\top}\Phi(u)}{W}\left(\mathbb{I} - \frac{\hat{F}\hat{F}^{\top}}{W}\right)X\Phi(v)\alpha$$

$$= \left[\frac{\hat{F}^{\top}\Phi(u)F}{W} - \frac{\hat{F}^{\top}\Phi(u)\hat{F}}{W}\frac{\hat{F}^{\top}F}{W}\right]\Lambda_0^{\top}\Phi(v)\alpha + \left[\frac{\hat{F}^{\top}\Phi(u)}{W} - \frac{\hat{F}^{\top}\Phi(u)\hat{F}}{W}\frac{\hat{F}^{\top}}{W}\right]\varepsilon\Phi(v)\alpha$$

$$= \hat{C}_a(u,v) + \hat{C}_b(u,v).$$

Similar to Sec.2, we decompose the complex-valued empirical process at the matrix level into the form of "signal" and "noise". Under the null hypothesis $\mathbb{H}_0$, (2) represents a non-anomalous factor model. Based on this, we consider $\hat{C}_a(u, v)$ and proceed to decompose it as follows:

$$\hat{C}_a(u,v) = \left[\frac{\hat{F}^{\top}\Phi(u)F}{W} - \frac{\hat{F}^{\top}\Phi(u)\hat{F}}{W}\frac{\hat{F}^{\top}F}{W}\right]\Lambda_0^{\top}\Phi(v)\alpha,$$

it represents the error between the estimated common factor $\hat{F}$ and the true factors F. The difference between the common factor and the residuals is represented as $F - \hat{F}A^{-1}$. This operation is a standard factor model error decomposition, we have

$$\hat{C}_a(u,v) = \left[\frac{\hat{F}^{\top}\Phi(u)(F - \hat{F}A^{-1})}{W} - \frac{\hat{F}^{\top}\Phi(u)\hat{F}}{W}\frac{\hat{F}^{\top}(F - \hat{F}A^{-1})}{W}\right]\Lambda_0^{\top}\Phi(v)\alpha$$

$$= \left[\frac{\hat{F}^{\top}\Phi(u)\hat{F}}{W}\frac{\hat{F}^{\top}(\hat{F} - FA)}{W} - \frac{\hat{F}^{\top}\Phi(u)(\hat{F} - FA)}{W}\right]A^{-1}\Lambda_0^{\top}\Phi(v)\alpha$$

$$= [\hat{C}_{aa}(u,v) - \hat{C}_{ab}(u,v)]A^{-1}\Lambda_0^{\top}\Phi(v)\alpha,$$

the asymptotically dominant term of $\hat{C}_a(u, v)$ is $\hat{C}_{ab}(u, v)$. Next, we focus on $\hat{C}_{ab}(u, v)$. Under the null hypothesis $\mathbb{H}_0$, we have the expression: $\hat{F} - FA = [\frac{\varepsilon\varepsilon^{\top}\hat{F}}{HW} + \frac{F\Lambda_0^{\top}\varepsilon^{\top}\hat{F}}{HW} + \frac{\varepsilon\Lambda_0 F^{\top}\hat{F}}{HW}]B_{HW}^{-1}$, it represents the error term between the factor loadings and the true factors, and is modeled through the residuals ε and the matrix Λ_0, then we have

$$-\hat{C}_{ab}(u,v)A^{-1}\Lambda_0^{\top}\Phi(v)\alpha = \frac{\hat{F}^{\top}\Phi(u)(\hat{F} - FA)}{W}A^{-1}\Lambda_0^{\top}\Phi(v)\alpha$$

$$= -\left[\frac{\hat{F}^{\top}\Phi(u)}{W}\frac{(\varepsilon + F\Lambda_0^{\top})\varepsilon^{\top}\hat{F}}{HW} + \frac{\hat{F}^{\top}\Phi(u)}{W}\frac{\varepsilon\Lambda_0 F^{\top}\hat{F}}{HW}\right]B_{HW}^{-1}A^{-1}\Lambda_0^{\top}\Phi(v)\alpha$$

$$= -[\hat{C}_{aba}(u,v) + \hat{C}_{abb}(u,v)+]B_{HW}^{-1}A^{-1}\Lambda_0^{\top}\Phi(v)\alpha.$$

By comparing the orders of the decomposition terms, we identify the asymptotically dominant term: $\hat{C}_{abb}(u,v)B_{HW}^{-1}A^{-1}\Lambda_0^{\top}\Phi(v)\alpha = \frac{\hat{F}^{\top}\Phi(u)\varepsilon\Lambda_0}{HW}\left(\frac{\Lambda_0^{\top}\Lambda_0}{H}\right)^{-1}\Lambda_0^{\top}\Phi(v)\alpha$ according to the definition of A. Now, we proceed to consider $\hat{C}_b(u, v)$.

$$\hat{C}_b(u,v) = \frac{\hat{F}^{\top}\Phi(u)\varepsilon\Phi(v)}{W} - \frac{\hat{F}^{\top}\Phi(u)\hat{F}}{W}\frac{\hat{F}^{\top}\varepsilon\Phi(v)}{W}$$

$$= \hat{C}_{ba}(u,v) - \hat{C}_{bb}(u,v),$$

we obtain the asymptotic dominant term $\hat{C}_{ba}(u,v)$. By the way, we combine $\hat{C}_{abb}(u,v)B_{HW}^{-1}A^{-1}\Lambda_0^\top\Phi(v)\alpha$ and $\hat{C}_{ba}(u,v)$ and have

$$
\hat{C}_{ba}(u,v) - \hat{C}_{abb}(u,v)B_{HW}^{-1}A^{-1}\Lambda_0^\top\Phi(v)\alpha = \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}\hat{F}_w\varepsilon_{hw}B\varphi(u,v)
$$

$$
= \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}(\hat{F}_w - A^\top F_w)\varepsilon_{hw}B\varphi(u,v) + \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}A^\top F_w\varepsilon_{hw}B\varphi(u,v),
$$

where $B = \lim_{H,W\to\infty}\hat{B}$, the asymptotic leading term of $\hat{C}_{ba}(u,v) - \hat{C}_{abc}(u,v)B_{HW}^{-1}A^{-1}\Lambda_0^\top\Phi(v)\alpha$ is given by: $\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}A^\top F_w\varepsilon_{hw}B\varphi(u,v)$. We can therefore conclude that

$$
\sup_{u,v\in\delta}\left\|\sqrt{HW}\hat{C}(u,v) - \frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}A^\top F_w\varepsilon_{hw}\varphi(u,v)B\right\| = o_p(1).
$$

It represents the difference between $\sqrt{HW}\hat{C}(u,v)$ and the second term, which converges at a rate of $o_p(1)$ in the large sample limit, consistent with the weak convergence theory of the complex-valued empirical process. Let $A_0 = \text{plim}_{H,W\to\infty}A$, and define $\hat{G}(u,v) = A_0^\top\left[\frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\varphi(u,v)\right]$, it is straightforward to show that $\frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\varphi(u,v)$ is asymptotically tight in $\mathcal{G}(\delta)$, where $\mathcal{G}(\delta)$ denotes the space of complex-valued continuous functions on δ. For any $u_1,u_2,v_1,v_2\in\delta$, we apply the Lagrange mean value theorem to obtain:

$$
\sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\left[\varphi(u_1,v_1) - \varphi(u_2,v_2)\right] = \sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\left[\frac{\partial\varphi(u,v)}{\partial u}(\xi,\eta)(u_1-u_2) + \frac{\partial\varphi(u,v)}{\partial v}(\xi,\eta)(v_1-v_2)\right],
$$

where (ξ,η) is a point on the intermediate path of (u_1,v_1) and (u_2,v_2), then

$$
E\left\|\frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\varphi(u_1,v_1) - \frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\varphi(u_2,v_2)\right\|^2
$$

$$
= \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}E\|F_wB\varepsilon_{hw}\|^2\left(\frac{\partial\varphi(u,v)}{\partial u}(\xi,\eta)(u_1-u_2) + \frac{\partial\varphi(u,v)}{\partial v}(\xi,\eta)(v_1-v_2)\right)^2.
$$

By the assumption that the residual terms form a Martingale Difference Sequence and applying the Cauchy-Schwarz inequality, we can derive the following:

$$
\frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W}E\|F_wB\varepsilon_{hw}\|^2 \leq \max\left(E\|F_w\|^4\right)^{\frac{1}{2}}\max\left(E\|\varepsilon_{hw}\|^4\right)^{\frac{1}{2}}\max\|B\|^2,
$$

where the order of last term is $O(1)$. Thus,

$$
E\left\|\frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\varphi(u_1,v_1) - \frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\varphi(u_2,v_2)\right\|^2 \leq C\|(u_1,v_1) - (u_2,v_2)\|^2,
$$

therefore, $\frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\varphi(u,v)$ is stochastically equicontinuous (Newey, 1991). It ensures the uniform behavior of the complex-valued empirical process over the space δ, as the sample sizes H and W increase, the process does not exhibit large fluctuations at different points, thus providing essential control conditions for proving the weak convergence and tightness of the process. Next, let $\ell(u,v) = BF_w\varepsilon_{hw}d^\top\varphi(u,v)$, where d is an arbitrary nonrandom vector with $\|d\| = 1$. Then, we have: $d^\top\left[\frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}F_wB\varepsilon_{hw}\varphi(u,v)\right] = \frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}\ell(u,v)$. We assume that $E[\ell(u,v)] = 0$. For some $r \geq 2$, applying the Cauchy-Schwarz inequality we also have the following bound:

$$
E[|\ell(u,v)|^r] \leq E[|B\varepsilon_{hw}|^{2r}]^{\frac{1}{2}}E[|d^\top F_w\varphi(u,v)|^{2r}]^{\frac{1}{2}} < \infty.
$$

This allows us to handle the higher-order moments and ensure their existence. Let $\mathcal{H}(u,v) =$ $\mathrm{Var}\left[\frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}\ell(u,v)\right] > 0$ for each fixed u, v. Then, according to Central Limit Theorem, we have:

$$\frac{1}{\sqrt{HW}\cdot\sqrt{\mathcal{H}(u,v)}}\sum_{h=1}^{H}\sum_{w=1}^{W}\ell(u,v) \xrightarrow{d} N(0,I).$$

Based on the Cramer-Wold device,

$$\hat{G}(u,v) = \frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}A_0^{\top}F_w B\varepsilon_{hw}\varphi(u,v) \xrightarrow{d} N(0,\Sigma),$$

for each fixed u, v, where $\Sigma = \lim_{H,W\to\infty}\mathrm{Var}\left(\frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}A_0^{\top}F_w B\varepsilon_{hw}\right)\left[\left(1 - \int_0^1 e^{i2\pi u\tau}\,d\tau\right)\left(1 - \int_0^1 e^{i2\pi v\tau}\,d\tau\right)\right]$,

where τ follows the uniform distribution $U(0,1)$. Given that δ is totally bounded, we have: $\hat{G}(u,v) \Rightarrow G(u,v)$ in $\mathcal{G}$, where $G(u,v)$ is a joint mean-zero, complex-valued Gaussian process. Therefore, we conclude that: $\sqrt{HW}\hat{C}(u,v) \Rightarrow G(u,v)$ in $\mathcal{G}$ as $(H,W) \to \infty$ under $\mathbb{H}_0$. ∎

B.2. Proof of Theorem 3.2

Given Proposition 3.1 under $\mathbb{H}_0$

$$\hat{D} = \iint_R \left\|\hat{G}(u,v)\right\|^2 W(u,v)\mathrm{d}u\mathrm{d}v + o_p(1),$$

where $o_p(1)$ denotes a small term that converges to zero in probability. To establish weak convergence, we need to prove that $\hat{D}$ converges in distribution to D as $(H,W) \to \infty$, there exists a compact subset $\delta \in \mathbb{R}$ large enough such that $\iint_{\delta^c}\|G(u,v)\|^2 W(u,v)\mathrm{d}u\mathrm{d}v < \frac{\epsilon^2}{2}$. for any positive value ϵ, δ^c denotes the complement of δ in $\mathbb{R}$ and we have:

$$\hat{D} = \iint_\delta \left\|\hat{G}(u,v)\right\|^2 W(u,v)\,du\,dv + \iint_{\delta^c}\left\|\hat{G}(u,v)\right\|^2 W(u,v)\,du\,dv,$$

and similarly for D:

$$D = \iint_\delta \|G(u,v)\|^2 W(u,v)\,du\,dv + \iint_{\delta^c}\|G(u,v)\|^2 W(u,v)\,du\,dv.$$

Thus, we need to show: $\hat{D}_\delta = \iint_\delta \left\|\hat{G}(u,v)\right\|^2 W(u,v)\,du\,dv$ converges in distribution to $D_\delta = \iint_\delta \|G(u,v)\|^2 W(u,v)\,du\,dv$; and $\hat{D}_{\delta^c} = \iint_{\delta^c}\left\|\hat{G}(u,v)\right\|^2 W(u,v)\,du\,dv$ converges in distribution to $D_{\delta^c} = \iint_{\delta^c}\|G(u,v)\|^2 W(u,v)\,du\,dv$.

We start by analyzing the part over δ:

$$\hat{D}_\delta = \iint_\delta \left\|\hat{G}(u,v)\right\|^2 W(u,v)\,du\,dv.$$

From Proposition 3.1, we know that $\hat{G}(u,v)$ converges to $G(u,v)$ in distribution as $(H,W) \to \infty$, i.e. $\hat{G}(u,v) \xrightarrow{d} G(u,v)$. To show that $\hat{D}_\delta$ converges to D_δ, we need to ensure that the integrals of $\hat{G}(u,v)$ and $G(u,v)$ converge, and that the integrands remain uniformly integrable. For any fixed u, v, we assume $\hat{G}(u,v)$ and $G(u,v)$ are integrable. By using Jensen's inequality and Fubini's theorem, we can bound the expectation For some $r \geq 2$:

$$E\left\|\hat{G}(u,v)\right\|^r \leq \|A_0\|^r E\left\|\frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W}F_w B\varepsilon_{hw}\right\|^r \left(\sup_{u,v\in\delta}\max|\varphi(u,v)|\right)^r < C.$$

This inequality ensures that $\hat{G}(u,v)$ is uniformly integrable over δ. With the dominated convergence theorem, we can then show that:

$$\iint_\delta \left\|\hat{G}(u,v)\right\|^2 W(u,v)\,du\,dv \xrightarrow{d} \iint_\delta \|G(u,v)\|^2 W(u,v)\,du\,dv.$$

By the dominated convergence theorem, we can establish that: $\hat{D}_\delta \xrightarrow{d} D_\delta$. Next, we consider the part over δ^c:

$$\hat{D}_{\delta^c} = \iint_{\delta^c} \left\|\hat{G}(u,v)\right\|^2 W(u,v)\,du\,dv.$$

We have that: $E\left(\iint_{\delta^c} \left\|\hat{G}(u,v)\right\|^2 W(u,v)\,du\,dv\right) \to E\left(\iint_{\delta^c} \|G(u,v)\|^2 W(u,v)\,du\,dv\right)$. By uniform integrability,

we know: $E\left(\iint_{\delta^c} \left\|\hat{G}(u,v)\right\|^2 W(u,v)\,du\,dv\right) \to \iint_{\delta^c} E\,\|G(u,v)\|^2 W(u,v)\,du\,dv$. Therefore, $\hat{D}_{\delta^c}$ is also uniformly

integrable, and by applying the dominated convergence theorem again, we can conclude that: $\hat{D}_{\delta^c} \xrightarrow{d} D_{\delta^c}$. By combining the results from $\hat{D}_\delta$ and $\hat{D}_{\delta^c}$, we get:

$$\hat{D} = \hat{D}_\delta + \hat{D}_{\delta^c} \xrightarrow{d} D_\delta + D_{\delta^c} = D.$$

Thus, $\hat{D}$ converges in distribution to D, i.e. $\hat{D} \xrightarrow{d} D$. ∎

B.3. Proof of Proposition 3.3

Consider the singular value decomposition (SVD) of $\mathbf{\Lambda}$: $\mathbf{\Lambda} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ where $\mathbf{U}$ and $\mathbf{V}$ are unitary matrices containing the left- and right-singular vectors of $\mathbf{\Lambda}$, respectively. $\mathbf{S}$ is an $H \times W$ diagonal matrix with singular values $s_{HW,1}, \ldots, s_{HW,L}$ along the diagonal, arranged in descending order, where $L = rank(S)$. Here, we assume that the first R singular values exert the most significant influence on the features, thereby preserving the maximum amount of information.

$$\mathbf{\Lambda} = \sum_{r=1}^{R} s_{HW,R} \mathbf{U}_R \mathbf{V}_R^\top + \sum_{l=R+1}^{L} s_{HW,l} \mathbf{U}_l \mathbf{V}_l^\top = \Lambda^{(R)} + \Lambda^{(-R)}.$$

Then we have

$$X = \mathbf{F}\Lambda^{(R)\top} + \mathbf{F}\Lambda^{(-R)\top} + \varepsilon.$$

This implies that the factor loading matrix is effectively represented in a R-dimensional space, where the asymptotic leading terms of $\mathbf{\Lambda}$ can be captured by the first R singular values. Given $\mathbf{A} = H^{-1}W^{-1}(\mathbf{\Lambda}^\top\mathbf{\Lambda})(\mathbf{F}^\top\hat{F})B_{HW}^{-1}$, under $\mathbb{H}_A$, $X = \mathbf{F}\mathbf{\Lambda}^\top + \varepsilon$. We can decompose $\hat{C}(u,v)$ like what we do in Proposition 3.1:

$$
\begin{aligned}
\hat{C}(u,v) &= \frac{\hat{F}^\top \Phi(u)}{W}\left(\mathbb{I} - \frac{\hat{F}\hat{F}^\top}{W}\right) X \Phi(v)\alpha \\
&= \frac{\hat{F}^\top \Phi(u)}{W}\left(\mathbb{I} - \frac{\hat{F}\hat{F}^\top}{W}\right) \mathbf{F}\mathbf{\Lambda}^\top \Phi(v)\alpha + \frac{\hat{F}^\top \Phi(u)}{W}\left(\mathbb{I} - \frac{\hat{F}\hat{F}^\top}{W}\right) \varepsilon \Phi(v)\alpha \\
&= \hat{C}_a(u,v) + \hat{C}_b(u,v).
\end{aligned}
$$

Similarly, when $K < R$, we obtain $\hat{C}_a(u,v)$ is the asymptotic leading term, we can decompose $\hat{C}_a(u,v)$ as:

$$
\begin{aligned}
\hat{C}_a(u,v) &= \hat{C}_{aa}(u,v) + \left(\frac{(\mathbf{FA})^\top \Phi(u)}{W}\right)\left[\mathbb{I} - \frac{\mathbf{FA}(\mathbf{FA})^\top}{W}\right] \mathbf{F}\mathbf{\Lambda}^\top \Phi(v)\alpha \\
&= \hat{C}_{aa}(u,v) + \hat{C}_{ab}(u,v).
\end{aligned}
$$

By comparing the order of each decomposed term, we found that $\hat{C}_{ab}(u,v)$ serves as the asymptotic leading term. As we have mentioned $\mathbf{\Lambda} = \Lambda^{(R)} + \Lambda^{(-R)}$, we have

$$
\begin{aligned}
\mathbf{A} &= \frac{\Lambda^{(R)\top}\Lambda^{(R)}\mathbf{F}^\top\hat{F}}{HW}B_{HW}^{-1} + \frac{\Lambda^{(-R)\top}\Lambda^{(-R)}\mathbf{F}^\top\hat{F}}{HW}B_{HW}^{-1} \\
&= \mathbf{A}_a + \mathbf{A}_b.
\end{aligned}
$$

Then

$$
\begin{aligned}
\hat{C}_{ab}(u,v) =& \left(\frac{(\mathbf{FA})^\top \Phi(u)}{W} \right) \left[\mathbb{I} - \frac{\mathbf{FA}(\mathbf{FA})^\top}{W} \right] \mathbf{F}\mathbf{\Lambda}^\top \Phi(v)\alpha \\
=& \left(\frac{(\mathbf{FA}_a)^\top \Phi(u)}{W} \right) \left[\mathbb{I} - \frac{\mathbf{FA}_a(\mathbf{FA}_a)^\top}{W} \right] \mathbf{F}\mathbf{\Lambda}^\top \Phi(v)\alpha \\
& - \left(\frac{(\mathbf{FA}_b)^\top \Phi(u)}{W} \right) \left[\frac{[(\mathbf{FA}_a)^\top + (\mathbf{FA}_b)^\top][\mathbf{FA}_a + \mathbf{FA}_b]}{W} - \mathbb{I} \right] \mathbf{F}\mathbf{\Lambda}^\top \Phi(v)\alpha \\
& - \left(\frac{(\mathbf{FA}_a)^\top \Phi(u)}{W} \right) \left[\frac{(\mathbf{FA}_a)^\top \mathbf{FA}_b + (\mathbf{FA}_b)^\top \mathbf{FA}_a + (\mathbf{FA}_b)^\top \mathbf{FA}_b}{W} \right] \mathbf{F}\mathbf{\Lambda}^\top \Phi(v)\alpha \\
=& \hat{C}_{aba}(u,v) - \hat{C}_{abb}(u,v) - \hat{C}_{abc}(u,v).
\end{aligned}
$$

By comparing the order of each decomposed term, the $\hat{C}_{aba}(u,v)$ is the asymptotic leading term of $\hat{C}_{ab}(u,v)$. The same operation can be applied:

$$
\begin{aligned}
\hat{C}_{aba}(u,v) =& B_{HW}^{-1} \frac{\hat{F}^\top \mathbf{F}\mathbf{\Lambda}^{(R)}{}^\top \mathbf{\Lambda}^{(R)} \mathbf{F}^\top \Phi(u) \mathbf{F}\mathbf{\Lambda}^\top \Phi(v)\alpha}{HW^2} \\
& - B_{HW}^{-1} \frac{\hat{F}^\top \mathbf{F}\mathbf{\Lambda}^{(R)}{}^\top \mathbf{\Lambda}^{(R)} \mathbf{F}^\top \Phi(u) \mathbf{F}\mathbf{\Lambda}^{(R)}{}^\top \mathbf{\Lambda}^{(R)} \mathbf{F}^\top \hat{F}}{H^2W^3} B_{HW}^{-2} \frac{\hat{F}^\top \mathbf{F}\mathbf{\Lambda}^{(R)}{}^\top \mathbf{\Lambda}^{(R)} \mathbf{F}^\top \mathbf{F}\mathbf{\Lambda}^\top \Phi(v)\alpha}{HW^2}.
\end{aligned}
$$

By spilting $\mathbf{\Lambda} = \mathbf{\Lambda}^{(R)} + \mathbf{\Lambda}^{(-R)}$, we have $\hat{C}_{aba}(u,v) = \hat{C}_{abaa}(u,v) + \hat{C}_{abab}(u,v)$. Obviously $\hat{C}_{abaa}(u,v)$ is the asymptotic leading term of $\hat{C}_{aba}(u,v)$.

$$
\begin{aligned}
\hat{C}_{abaa}(u,v) =& B_{HW}^{-1} \frac{\hat{F}^\top \mathbf{F}\mathbf{V}^{(R)}\mathbf{S}^{(R)^2}\mathbf{V}^{(R)^\top} \mathbf{F}^\top \Phi(u) \mathbf{F}\mathbf{V}^{(R)}\mathbf{S}^{(R)}\mathbf{U}^{(R)}\Phi(v)\alpha}{HW^2} \\
& - B_{HW}^{-1} \frac{\hat{F}^\top \mathbf{F}\mathbf{V}^{(R)}\mathbf{S}^{(R)^2}\mathbf{V}^{(R)^\top} \mathbf{F}^\top \Phi(u) \mathbf{F}\mathbf{V}^{(R)}\mathbf{S}^{(R)^2}\mathbf{V}^{(R)^\top} \mathbf{F}^\top \hat{F}}{H^2W^3} \\
& \times B_{HW}^{-2} \frac{\hat{F}^\top \mathbf{F}\mathbf{V}^{(R)}\mathbf{S}^{(R)^2}\mathbf{V}^{(R)^\top} \mathbf{F}^\top \mathbf{F}\mathbf{V}^{(R)}\mathbf{S}^{(R)}\mathbf{U}^{(R)}\Phi(v)\alpha}{HW^2} \\
\to \hat{C}_{abaa}(u,v) \propto& (B_{HW}^{-\frac{1}{2}})(\frac{1}{HW}\mathbf{S}^{(R)}\mathbf{V}^{(R)^\top}\mathbf{F}^\top \Phi(u)\mathbf{F}\mathbf{V}^{(R)}\mathbf{S}^{(R)})(\sqrt{H}\mathbf{U}^{(R)^\top}\Phi(v)\alpha) + o_p(1).
\end{aligned}
$$

Each decomposition term in matrix $\hat{C}_{abaa}(u,v)$ has a nonzero order. , thus $\hat{C}_{abaa}(u,v) \Rightarrow \mathcal{C}(u,v)$ is a non-zero spectrum.■

B.4. Proof of Theorem 3.4

When $K < R$, $\mathcal{C}(u,v)$ is a nonzero spectrum. In the vast majority of cases, $C(u,v)$ is non-zero and meaningful, thus ensuring that $\mathcal{C}(u,v) > c_{HW}$ holds true in the limit. As $(H,W) \to \infty$, $\hat{D} = HW \iint_{\mathbb{R}} \|\hat{C}(u,v)\|^2 W(u,v)\, du\, dv$. and $c_{HW} \to 0$ Therefore, for the majority of samples, $\hat{D}$ will inevitably be greater than a small c_{HW} This leads to the probability $P(\hat{D} > c_{HW}) \to 1$.■

B.5. Proof of Proposition 3.5

Let $g^\dagger$ is a set of g_{hw}, Then under $\mathbb{H}_A(\tau_{HW})$, $\varepsilon^\dagger = \varepsilon + \tau_{HW}g_{hw}$, and $X = F\Lambda_0^\top + \varepsilon^\dagger$, then

$$
\begin{aligned}
\hat{C}(u,v) =& \frac{\hat{F}^\top \Phi(u)}{W}\left(\mathbb{I} - \frac{\hat{F}\hat{F}^\top}{W} \right) F\Lambda_0^\top \Phi(v)\alpha + \frac{\hat{F}^\top \Phi(u)}{W}\left(\mathbb{I} - \frac{\hat{F}\hat{F}^\top}{W} \right)\varepsilon^\dagger \Phi(v)\alpha \\
=& \hat{C}'_a(u,v) + \hat{C}'_b(u,v).
\end{aligned}
$$

17

We can follow the procedure outlined in Proposition 3.1 and decompose $\hat{C}'_a(u,v)$.

$$\hat{C}'_a(u,v) = \left[\frac{\hat{F}^\top \Phi(u)(F - \hat{F}A^{-1})}{W} - \frac{\hat{F}^\top \Phi(u)\hat{F}}{W}\frac{\hat{F}^\top(F - \hat{F}A^{-1})}{W}\right] \Lambda_0^\top \Phi(v)\alpha$$

$$= \left[\frac{\hat{F}^\top \Phi(u)\hat{F}}{W}\frac{\hat{F}^\top(\hat{F} - FA)}{W} - \frac{\hat{F}^\top \Phi(u)(\hat{F} - FA)}{W}\right] A^{-1}\Lambda_0^\top \Phi(v)\alpha$$

$$= [\hat{C}'_{aa}(u,v) - \hat{C}'_{ab}(u,v)]A^{-1}\Lambda_0^\top \Phi(v)\alpha.$$

By comparing the order of each term, we find that $\hat{C}'_{ab}(u,v)$ is the asymptotic leading term of $\hat{C}'_a(u,v)$.

$$\hat{F} - FA = \left[\frac{\varepsilon^\dagger \varepsilon^{\dagger\top}\hat{F}}{HW} + \frac{F\Lambda_0^\top \varepsilon^{\dagger\top}\hat{F}}{HW} + \frac{\varepsilon^\dagger \Lambda_0 F^\top \hat{F}}{HW}\right] B_{HW}^{-1}.$$

Thus,

$$-\hat{C}'_{ab}(u,v)A^{-1}\Lambda_0^\top \Phi(v)\alpha = \frac{\hat{F}^\top \Phi(u)(\hat{F} - FA)}{W}A^{-1}\Lambda_0^\top \Phi(v)\alpha$$

$$= -\left[\frac{\hat{F}^\top \Phi(u)}{W}\frac{\varepsilon^{\dagger\top}\hat{F}(\varepsilon^\dagger + F\Lambda_0^\top)}{HW} + \frac{\hat{F}\Phi(u)}{W}\frac{\varepsilon^\dagger \Lambda_0 F^\top \hat{F}}{HW}\right] B_{HW}^{-1}A^{-1}\Lambda_0^\top \Phi(v)\alpha$$

$$= -\left[\hat{C}'_{aba}(u,v) + \hat{C}'_{abb}(u,v)\right] B_{HW}^{-1}A^{-1}\Lambda_0^\top \Phi(v)\alpha.$$

By the same way, the $\hat{C}'_{abb}(u,v)$ is the asymptotic leading term of $\hat{C}'_{ab}(u,v)$. We have

$$\hat{C}'_{abb}(u,v)B_{HW}^{-1}A^{-1}\Lambda_0^\top \Phi(v)\alpha = \frac{(\hat{F} - AF)^\top \Phi(u)\varepsilon^\dagger \Lambda_0}{HW}\left(\frac{\Lambda_0^\top \Lambda_0}{H}\right)^{-1}\Lambda_0^\top \Phi(v)\alpha + \frac{A^\top F^\top \Phi(u)\varepsilon^\dagger \Lambda_0}{HW}\left(\frac{\Lambda_0^\top \Lambda_0}{H}\right)^{-1}\Lambda_0^\top \Phi(v)\alpha$$

$$= \hat{C}'_{abba}(u,v) + \hat{C}'_{abbb}(u,v),$$

and $\hat{C}'_{abbb}(u,v)$ is the asymptotic leading term. Next, we consider $\hat{C}'_b(u,v)$.

$$\hat{C}'_b(u,v) = \frac{\hat{F}^\top \Phi(u)}{W}\left(\mathbb{I} - \frac{\hat{F}\hat{F}^\top}{W}\right)\varepsilon^\dagger \Phi(v)\alpha$$

$$= \frac{\hat{F}^\top \Phi(u)\varepsilon^\dagger \Phi(v)\alpha}{W} - \frac{\hat{F}^\top \Phi(u)\hat{F}}{W}\frac{\hat{F}^\top \varepsilon^\dagger \Phi(v)\alpha}{W}$$

$$= \hat{C}'_{ba}(u,v) - \hat{C}'_{bb}(u,v).$$

By comparing the order, we obtain $\hat{C}'_{ba}(u,v)$ is the asymptotic leading term of $\hat{C}'_b(u,v)$, then we decompose $\hat{C}'_{ba}(u,v)$.

$$\hat{C}'_{ba}(u,v) = \frac{(\hat{F} - AF)^\top \Phi(u)\varepsilon^\dagger \Phi(v)\alpha}{W} + \frac{A^\top F^\top \Phi(u)\varepsilon^\dagger \Phi(v)\alpha}{W}$$

$$= \hat{C}'_{baa}(u,v) + \hat{C}'_{bab}(u,v),$$

and the asymptotic leading term is $\hat{C}'_{bab}(u,v)$, then we combine the asymptotic terms $\hat{C}'_{bab}(u,v)$ and $-\hat{C}'_{abbb}(u,v)$ and obtain

$$\hat{C}'_{bab}(u,v) - \hat{C}'_{abbb}(u,v) = \frac{A^\top F^\top \Phi(u)\varepsilon^\dagger \Phi(v)\alpha}{W} - \frac{A^\top F^\top \Phi(u)\varepsilon^\dagger \Lambda_0}{HW}\left(\frac{\Lambda_0^\top \Lambda_0}{H}\right)^{-1}\Lambda_0^\top \Phi(v)\alpha$$

$$= \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} A^\top F_w \varepsilon_{hw} B\varphi(u,v) + \frac{\tau_{HW}}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} A^\top F_w F_w^\top g_{hw} B\varphi(u,v).$$

We have

$$\sup_{u,v\in\delta}\left\|\sqrt{HW}\hat{C}(u,v) - \frac{1}{\sqrt{HW}}\sum_{h=1}^{H}\sum_{w=1}^{W} A_0^\top F_w \varepsilon_{hw} B\varphi(u,v) - \frac{1}{HW}\sum_{h=1}^{H}\sum_{w=1}^{W} A_0^\top F_w F_w^\top g_{hw} B\varphi(u,v)\right\| = o_p(1).$$

18

Let $\hat{G}(u,v) = \frac{1}{\sqrt{HW}} \sum_{h=1}^{H} \sum_{w=1}^{W} A_0^\top F_w \varepsilon_{hw} B\varphi(u,v)$ and $\hat{\phi}(u,v) = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} A_0^\top F_w F_w^\top g_{hw} B\varphi(u,v)$. As we have mentioned in Proposition 3.1 that $\hat{G}(u,v) \Rightarrow G(u,v)$, Such that

$$\hat{\phi}(u,v) = \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} A_0^\top F_w F_w^\top g_{hw} B\varphi(u,v)$$

$$= \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} A_0^\top (F_w F_w^\top - \psi_F) g_{hw} B\varphi(u,v) + \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} A_0^\top \psi_F g_{hw} B\varphi(u,v).$$

We have $H^{-1}W^{-1} \sum_{h=1}^{H} \sum_{w=1}^{W} g_{hw} = 0$ it is straightforward to show that $E\left[\frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} A_0^\top (F_w F_w^\top - \psi_F) g_{hw} B\varphi(u,v)\right] = 0$, and $E\left\| \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} A_0^\top (F_w F_w^\top - \psi_F) g_{hw} B\varphi(u,v) \right\|^2 = O(H^{-1}W^{-1})$. Thus,

$$\sup_{u,v \in \delta} \left\| \frac{1}{HW} \sum_{h=1}^{H} \sum_{w=1}^{W} A_0^\top (F_w F_w^\top - \psi_F) g_{hw} B\varphi(u,v) \right\| = O_p(H^{-1/2}W^{-1/2}).$$

It follows that $\hat{\phi}(u,v) \Rightarrow \phi(u,v)$, where $\phi(u,v)$ denotes a pseudo-covariance between the Fourier series $e^{2\pi i\left(\frac{uh}{H} + \frac{vw}{W}\right)}$ and $g_{hw}B$. So we have $\sqrt{HW}\hat{C}(u,v) \Rightarrow \phi(u,v) + G(u,v)$ under $\mathbb{H}_A(\tau_{HW})$ with $\tau_{HW} = H^{-1/2}W^{-1/2}$. ∎

B.6. Proof of Theorem 3.6

Under $\mathbb{H}_A(\tau_{HW}): \mathbb{H}_A(\tau_{HW}) = \lambda_{hw} = \lambda_0 + \tau_{HW} g_{hw}$. By Proposition 3.5, we have

$$\sqrt{HW}\hat{C}(u,v) \Rightarrow \phi(u,v) + G(u,v).$$

By the continuous mapping theorem, we have

$$\iint_\delta \|\sqrt{HW}\hat{C}(u,v)\|^2 W(u,v)\, du\, dv \xrightarrow{d} \iint_\delta \|\phi(u,v) + G(u,v)\|^2 W(u,v)\, du\, dv.$$

By similar arguments in the Proof of Theorem 3.2, it follows

$$\hat{D} \xrightarrow{d} \iint_{\mathbb{R}} \|\phi(u,v) + G(u,v)\|^2 W(u,v)\, du\, dv$$

. ∎

C. The Division of Datasets

MVTec-AD Dataset: MVTec-AD dataset (Bergmann et al., 2019a) is collected for real-world industrial production scenarios unsupervised anomaly detection. There are 10 object types: Bottle, Cable, Capsule, Hazelnut, Metal Nut, Pill, Screw, Toothbrush, Transistor, Zipper. And 5 texture types: Carpet, Grid, Leather, Tile, Wood. The dataset has totaling 5,354 images with pixel-level annotations. Among them, the training set contains 3,629 anomaly-free images, and the testing set contains 1,725 images with both normal and anomaly samples.

VisA Dataset: VisA dataset (Zou et al., 2022) is a large dataset collected for industrial anomaly detection, covering 12 objects in 3 domains: Complex structure, Multiple instances and Single instance. Among them, Complex structure domain has 4 objects: PCB1, PCB2, PCB3, PCB4. Multiple instances domain has 4 objects: Macaroni 1, Macaroni 2, Capsules, Candles. Single instance domain has 4 objects: Cashew, Chewing gum, Fryum, Pipe fryum. The dataset has 10,821 images with pixel-level annotations. Among them, the training set contains 8,721 anomaly-free images, and the testing set contains 2,100 images with both normal and anomaly samples.

D. Hyperparameter Selection Study Results

D.1. Hyperparameter Selection Settings

Equation 7 defined the decentralization strategy. This section reports the results of different sampling functions of mask sampling threshold s, different mean of mask sampling threshold s, as well as the number of sampling epochs N. We first

conduct experiments with standard uniform ($U(0, 1)$) and standard normal distribution ($N(0, 1)$) sampling functions with the same setting of μ_s on MvTec-AD dataset, where μ_s changes from 0.1 to 0.9 increasing by 0.1 at a time, and the number of sampling epochs N are both set to 5 by default. Then we further conduct experiments for different μ_s on VisA dataset, as well as different N on both MvTec-AD and VisA dataset where sampling function is set to uniform distribution by default. When exploring one of the two hyperparameters, the other is set to the default, the settings are as follows: the μ_s changes from 0.1 to 0.9 increasing by 0.1 at a time, and the number of sampling epochs N changes from 5 to 35, increasing by 5 at a time, the default μ_s is 0.5, and the default N is 5. Meanwhile, we conducted parameter selection experiments for different regularization coefficients α

D.2. Detailed Results

Table 5. Detailed anomaly detection/localization results with Img-AUROC / Pix-AUROC / PRO (in %) metric of different μ_s on MvTec-AD under sampling function with Uniform Distribution.

Category	$\mu_s = 0.1$	$\mu_s = 0.2$	$\mu_s = 0.3$	$\mu_s = 0.4$	$\mu_s = 0.5$	$\mu_s = 0.6$	$\mu_s = 0.7$	$\mu_s = 0.8$	$\mu_s = 0.9$
carpet	97.99 / 99.59 / 97.46	99.16 / 99.51 / 97.29	93.10 / 99.43 / 97.61	96.11 / 99.55 / 97.37	99.00 / 99.45 / 96.77	97.03 / 99.50 / 97.46	98.07 / 99.43 / 97.38	99.36 / 99.33 / 96.96	95.71 / 99.17 / 95.75
Grid	100.00 / 99.07 / 95.60	100.00 / 99.02 / 95.67	100.00 / 99.07 / 96.03	100.00 / 99.15 / 96.24	100.00 / 98.90 / 94.91	99.83 / 98.31 / 91.64	99.83 / 98.03 / 90.23	98.75 / 98.11 / 91.22	95.91 / 97.55 / 90.10
Leather	100.00 / 99.57 / 98.79	90.46 / 98.91 / 97.38	100.00 / 99.55 / 98.78	100.00 / 99.68 / 99.12	100.00 / 99.68 / 99.08	100.00 / 99.67 / 99.11	100.00 / 99.67 / 99.11	100.00 / 99.68 / 99.17	100.00 / 99.69 / 99.24
Tile	100.00 / 98.88 / 95.93	100.00 / 98.68 / 95.15	100.00 / 98.46 / 94.44	100.00 / 98.72 / 95.35	100.00 / 98.69 / 95.13	100.00 / 98.40 / 94.95	100.00 / 98.49 / 94.86	100.00 / 98.76 / 95.38	99.82 / 98.48 / 94.94
Wood	100.00 / 97.61 / 92.37	100.00 / 96.91 / 91.39	99.82 / 97.27 / 91.72	99.74 / 97.24 / 91.37	99.82 / 97.52 / 91.25	99.82 / 96.39 / 90.65	99.91 / 96.66 / 91.71	99.82 / 96.44 / 91.87	99.56 / 96.42 / 92.38
Pill	99.15 / 98.91 / 96.58	99.59 / 99.11 / 97.48	98.91 / 99.00 / 97.54	99.51 / 98.91 / 97.53	99.59 / 98.79 / 97.64	99.32 / 98.73 / 97.70	98.47 / 98.59 / 97.51	98.61 / 98.54 / 97.48	94.84 / 98.29 / 97.11
Transistor	98.62 / 94.75 / 82.43	99.29 / 94.73 / 84.15	99.33 / 95.27 / 84.55	99.67 / 94.41 / 84.11	99.42 / 95.32 / 84.98	98.92 / 96.10 / 86.33	98.88 / 96.06 / 85.79	98.62 / 96.67 / 86.20	98.67 / 95.67 / 85.63
Cable	95.80 / 96.14 / 85.23	97.40 / 97.67 / 88.63	98.16 / 98.31 / 90.36	98.52 / 98.50 / 91.46	99.03 / 98.62 / 91.10	99.46 / 98.54 / 92.00	98.80 / 98.60 / 91.49	98.67 / 98.79 / 92.78	97.62 / 98.76 / 92.59
Zipper	98.98 / 99.28 / 96.03	98.50 / 99.28 / 96.50	98.69 / 99.35 / 97.52	95.48 / 99.33 / 97.59	98.35 / 99.25 / 97.56	98.21 / 99.23 / 97.50	98.35 / 99.14 / 97.43	98.19 / 98.87 / 96.81	93.17 / 98.33 / 95.19
Toothbrush	100.00 / 98.23 / 89.54	99.44 / 98.51 / 91.85	99.72 / 98.48 / 91.19	100.00 / 99.03 / 94.45	99.17 / 99.05 / 93.62	100.00 / 99.23 / 95.01	100.00 / 99.21 / 94.98	100.00 / 99.23 / 94.92	100.00 / 99.24 / 94.81
Metal Nut	97.12 / 96.99 / 84.89	97.75 / 96.57 / 85.07	97.90 / 96.42 / 83.71	97.12 / 96.60 / 85.31	97.41 / 96.87 / 84.97	97.56 / 97.09 / 86.12	96.63 / 97.12 / 86.03	96.63 / 96.82 / 85.50	94.82 / 96.80 / 84.79
Hazelnut	99.93 / 98.99 / 93.92	98.32 / 99.07 / 93.71	99.89 / 99.25 / 94.39	98.82 / 99.05 / 93.74	98.57 / 99.10 / 93.95	98.82 / 99.00 / 93.73	98.82 / 99.11 / 93.79	99.14 / 99.09 / 93.52	98.71 / 99.25 / 94.44
Screw	85.00 / 97.96 / 91.58	92.38 / 98.11 / 91.62	92.21 / 98.55 / 94.36	93.93 / 98.91 / 95.69	96.02 / 99.23 / 96.69	94.32 / 99.24 / 96.74	94.63 / 99.13 / 96.33	90.72 / 98.98 / 95.47	87.93 / 98.72 / 94.40
Capsule	96.17 / 91.14 / 81.96	94.73 / 95.79 / 86.96	94.89 / 96.60 / 88.00	94.18 / 96.12 / 88.04	94.10 / 96.25 / 88.18	94.30 / 96.53 / 88.24	93.82 / 97.28 / 89.60	94.38 / 97.63 / 90.78	95.97 / 97.77 / 91.73
Bottle	99.84 / 98.03 / 92.71	99.92 / 97.87 / 92.11	99.68 / 97.87 / 92.18	99.68 / 97.69 / 91.61	99.92 / 98.05 / 92.75	99.92 / 97.97 / 92.72	100.00 / 97.82 / 92.40	100.00 / 97.82 / 92.20	100.00 / 97.78 / 91.90
Average	97.91 / 97.68 / 91.67	97.80 / 97.98 / 92.33	98.15 / 98.19 / 92.83	98.18 / 98.19 / 93.27	98.69 / 98.32 / 93.24	98.50 / 98.26 / 93.33	98.41 / 98.29 / 93.24	98.19 / 98.32 / 93.35	96.85 / 98.13 / 93.00

Table 6. Detailed anomaly detection/localization results with Img-AUROC / Pix-AUROC / PRO (in %) metric of different μ_s on MvTec-AD under sampling function with Normal Distribution.

Category	$\mu_s = 0.1$	$\mu_s = 0.2$	$\mu_s = 0.3$	$\mu_s = 0.4$	$\mu_s = 0.5$	$\mu_s = 0.6$	$\mu_s = 0.7$	$\mu_s = 0.8$	$\mu_s = 0.9$
carpet	99.52 / 99.53 / 97.13	97.03 / 99.44 / 96.92	95.47 / 99.19 / 95.63	98.48 / 99.43 / 96.91	99.40 / 99.49 / 96.89	96.79 / 99.43 / 97.01	94.82 / 99.30 / 96.62	89.04 / 98.53 / 94.41	72.79 / 95.65 / 84.69
Grid	100.00 / 99.04 / 96.36	99.92 / 99.11 / 96.39	99.83 / 99.14 / 96.68	100.00 / 98.98 / 95.58	100.00 / 98.43 / 93.16	99.50 / 97.35 / 86.77	98.66 / 97.35 / 87.03	98.83 / 97.57 / 87.38	98.58 / 97.65 / 90.84
Leather	100.00 / 99.64 / 98.94	100.00 / 99.69 / 99.20	100.00 / 99.68 / 99.23	100.00 / 99.68 / 99.19	100.00 / 99.68 / 99.18	100.00 / 99.68 / 99.20	100.00 / 99.68 / 99.24	100.00 / 99.66 / 99.22	99.93 / 99.52 / 99.01
Tile	93.25 / 96.99 / 93.78	100.00 / 98.51 / 94.92	99.86 / 98.70 / 95.66	100.00 / 98.68 / 95.22	100.00 / 98.64 / 95.29	100.00 / 98.40 / 95.36	100.00 / 97.80 / 93.50	99.68 / 98.26 / 94.88	99.93 / 96.67 / 91.75
Wood	100.00 / 97.35 / 91.15	100.00 / 97.49 / 91.81	99.56 / 96.18 / 90.74	99.74 / 96.76 / 91.05	99.30 / 96.67 / 91.30	100.00 / 96.64 / 91.76	100.00 / 96.39 / 91.45	100.00 / 96.21 / 91.42	98.25 / 95.34 / 90.10
Pill	95.36 / 98.72 / 95.91	99.32 / 98.78 / 96.49	99.97 / 98.89 / 97.26	99.65 / 98.78 / 97.58	99.18 / 98.68 / 97.77	97.98 / 98.56 / 97.36	98.17 / 98.51 / 97.37	92.20 / 98.54 / 97.36	97.49 / 98.48 / 97.12
Transistor	99.42 / 96.64 / 85.07	99.38 / 95.59 / 84.96	99.33 / 95.31 / 84.75	99.42 / 95.00 / 84.69	99.17 / 95.28 / 85.38	99.12 / 95.57 / 85.80	98.92 / 95.39 / 85.49	98.58 / 96.02 / 85.74	98.62 / 95.65 / 86.35
Cable	98.65 / 98.54 / 91.17	98.24 / 98.35 / 90.59	98.71 / 98.32 / 90.53	97.75 / 98.49 / 91.15	98.61 / 98.29 / 90.39	98.67 / 98.37 / 90.82	97.88 / 98.51 / 91.44	98.29 / 98.80 / 92.50	99.12 / 98.73 / 92.39
Zipper	95.67 / 99.32 / 97.24	98.77 / 99.33 / 97.57	95.61 / 99.28 / 97.74	95.77 / 99.17 / 97.50	95.56 / 99.10 / 97.35	95.64 / 98.85 / 96.95	93.78 / 98.46 / 95.69	91.65 / 98.00 / 94.43	92.99 / 98.18 / 95.00
Toothbrush	100.00 / 99.11 / 90.90	100.00 / 99.05 / 93.70	100.00 / 99.01 / 94.48	100.00 / 99.10 / 94.87	100.00 / 99.27 / 95.19	100.00 / 99.24 / 95.13	99.72 / 99.26 / 95.20	99.72 / 99.28 / 95.16	100.00 / 99.26 / 95.16
Metal Nut	96.68 / 96.46 / 83.19	96.73 / 97.27 / 85.01	97.02 / 96.15 / 84.26	97.56 / 96.90 / 85.53	97.31 / 96.89 / 85.79	95.85 / 96.62 / 85.08	96.87 / 96.23 / 84.49	96.73 / 96.38 / 84.54	94.48 / 96.41 / 83.76
Hazelnut	98.61 / 98.55 / 93.28	98.43 / 99.03 / 93.93	98.43 / 99.04 / 94.15	98.57 / 99.02 / 94.33	98.82 / 99.00 / 94.00	98.96 / 99.19 / 93.97	99.11 / 99.40 / 94.54	99.21 / 99.39 / 94.06	99.11 / 99.31 / 94.05
Screw	87.09 / 98.31 / 92.97	93.11 / 98.66 / 94.70	93.65 / 98.85 / 95.33	92.91 / 99.20 / 96.72	94.77 / 99.39 / 97.16	94.20 / 99.37 / 97.20	93.87 / 99.21 / 96.71	91.23 / 99.22 / 96.46	94.10 / 99.22 / 96.51
Capsule	96.13 / 95.67 / 88.72	97.25 / 96.13 / 87.62	94.89 / 95.24 / 87.64	93.18 / 96.88 / 88.69	93.94 / 96.39 / 88.20	94.65 / 97.10 / 89.07	93.82 / 96.22 / 88.42	93.74 / 97.14 / 89.56	94.26 / 97.12 / 89.59
Bottle	99.92 / 98.51 / 93.91	99.92 / 98.39 / 93.68	100.00 / 97.89 / 92.31	99.52 / 98.08 / 92.68	99.92 / 98.05 / 92.98	99.92 / 97.82 / 92.33	99.92 / 97.73 / 92.10	99.92 / 97.75 / 92.02	99.92 / 97.75 / 92.03
Average	97.35 / 98.16 / 92.65	98.54 / 98.32 / 93.17	98.16 / 98.06 / 93.09	98.17 / 98.28 / 93.45	98.40 / 98.22 / 93.34	98.09 / 98.15 / 92.92	97.70 / 97.96 / 92.62	96.59 / 98.05 / 92.61	95.97 / 97.66 / 91.89

(a) Uniform Distribution
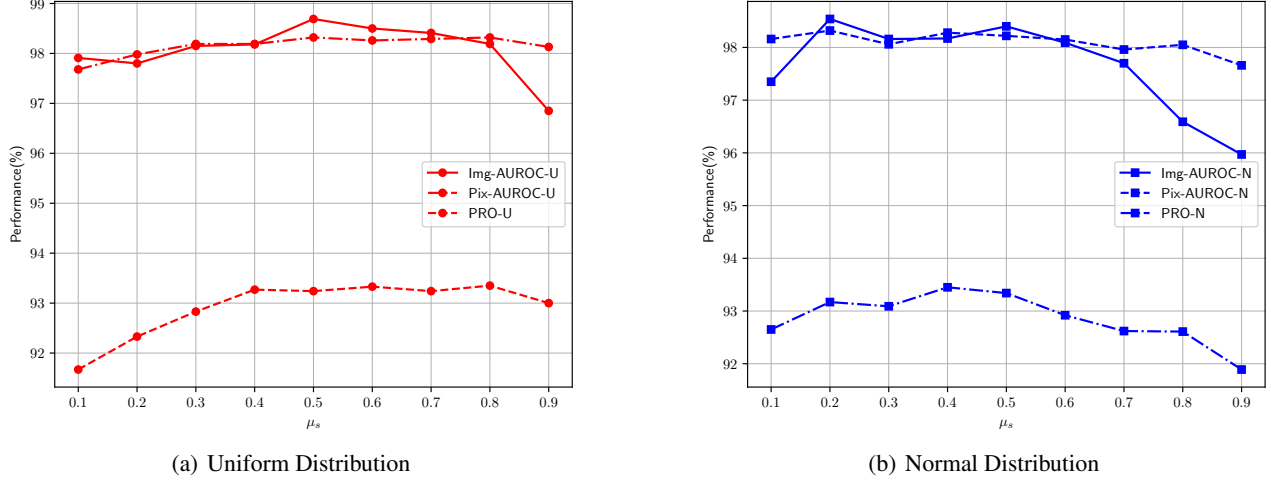
(b) Normal Distribution

Figure 5. Average performance of anomaly detection and localization of the mask sampling function with uniform distribution and normal distribution as the μ_s changes, U represents uniform distribution, N represents normal distribution.

Table 7. Detailed anomaly detection/localization results with Img-AUROC / Pix-AUROC / PRO (in %) metric on VisA under different μ_s.

Category	$\mu_s = 0.1$	$\mu_s = 0.2$	$s = 0.3$	$\mu_s = 0.4$	$\mu_s = 0.5$	$\mu_s = 0.6$	$\mu_s = 0.7$	$\mu_s = 0.8$	$\mu_s = 0.9$
PCB1	98.03 / 99.43 / 91.76	98.32 / 99.38 / 91.77	97.00 / 99.37 / 91.33	97.30 / 99.53 / 92.77	97.46 / 99.42 / 93.04	98.34 / 99.58 / 93.78	98.15 / 99.60 / 93.74	97.32 / 99.57 / 93.78	97.51 / 99.55 / 93.61
PCB2	99.43 / 98.61 / 86.74	99.38 / 98.59 / 86.52	99.53 / 98.46 / 84.49	99.17 / 98.73 / 88.74	99.22 / 98.70 / 87.33	98.60 / 98.88 / 90.30	99.01 / 99.03 / 91.27	99.16 / 99.01 / 91.54	98.50 / 98.98 / 90.97
PCB3	98.78 / 99.22 / 94.35	98.42 / 99.21 / 94.56	98.25 / 99.13 / 94.56	98.68 / 99.21 / 94.95	98.89 / 99.14 / 95.34	99.00 / 99.25 / 95.23	99.29 / 99.26 / 95.04	98.94 / 99.23 / 94.75	97.41 / 99.22 / 94.36
PCB4	99.53 / 96.23 / 84.98	99.36 / 93.21 / 77.71	99.27 / 94.45 / 79.99	99.31 / 93.62 / 77.98	99.28 / 94.59 / 80.87	99.06 / 96.37 / 82.61	99.24 / 97.54 / 85.67	99.42 / 97.67 / 86.61	99.72 / 98.08 / 87.78
Macaroni 1	96.50 / 99.73 / 98.57	96.75 / 99.72 / 98.37	97.58 / 99.74 / 98.59	97.84 / 99.72 / 98.41	96.94 / 99.78 / 98.82	97.51 / 99.80 / 98.83	95.36 / 99.68 / 97.48	96.02 / 99.66 / 96.89	94.38 / 99.59 / 96.27
Macaroni 2	91.44 / 99.78 / 99.16	88.22 / 99.67 / 98.91	89.85 / 99.74 / 98.99	91.14 / 99.79 / 99.16	90.08 / 99.74 / 98.81	93.76 / 99.82 / 99.24	87.62 / 99.81 / 99.14	84.61 / 99.83 / 99.11	83.47 / 99.81 / 99.04
Capsules	96.22 / 99.69 / 98.00	95.50 / 99.77 / 98.01	96.37 / 99.82 / 97.97	97.13 / 99.85 / 98.46	95.92 / 99.87 / 98.19	94.80 / 99.87 / 98.35	93.50 / 99.85 / 98.59	92.72 / 99.87 / 98.30	92.57 / 99.84 / 97.48
Candle	94.67 / 94.68 / 87.65	93.49 / 94.72 / 87.08	89.55 / 95.96 / 87.18	88.95 / 96.28 / 88.47	88.44 / 95.84 / 89.20	85.71 / 95.28 / 88.93	88.52 / 95.85 / 89.23	89.08 / 95.74 / 88.53	88.22 / 95.72 / 88.68
Cashew	97.28 / 97.62 / 83.92	98.00 / 98.93 / 85.69	97.06 / 98.71 / 84.51	97.40 / 99.15 / 88.24	96.04 / 98.80 / 87.36	94.62 / 98.93 / 88.92	97.88 / 99.18 / 88.26	99.28 / 99.21 / 90.10	97.96 / 99.11 / 89.39
Chewinggum	97.54 / 97.82 / 76.77	98.08 / 97.70 / 76.34	97.36 / 98.10 / 78.29	97.40 / 98.11 / 78.79	97.28 / 98.29 / 80.10	97.52 / 98.45 / 81.98	97.78 / 98.46 / 84.03	97.20 / 98.49 / 83.72	96.02 / 98.33 / 82.81
Fryum	97.96 / 95.34 / 88.91	99.04 / 94.30 / 88.84	99.62 / 94.26 / 88.63	98.92 / 94.95 / 88.54	97.84 / 94.71 / 88.93	98.32 / 94.68 / 88.56	98.12 / 95.76 / 89.20	97.18 / 96.42 / 89.96	90.04 / 96.81 / 91.65
Pipe fryum	97.66 / 99.31 / 94.91	97.62 / 99.35 / 94.94	98.12 / 99.35 / 94.96	96.88 / 99.39 / 95.48	98.68 / 99.41 / 95.82	98.20 / 99.39 / 95.97	98.18 / 99.46 / 95.75	97.58 / 99.43 / 96.03	96.36 / 99.46 / 95.80
Average	97.09 / 98.12 / 90.48	96.85 / 97.88 / 89.88	96.63 / 98.09 / 89.96	96.68 / 98.19 / 90.83	96.34 / 98.19 / 91.15	96.29 / 98.36 / 91.89	96.05 / 98.62 / 92.28	95.71 / 98.68 / 92.44	94.35 / 98.71 / 92.32

Table 8. Detailed anomaly detection/localization results with Img-AUROC / Pix-AUROC / PRO (in %) metric on MvTec-AD under different N.

Category	$N = 5$	$N = 10$	$N = 15$	$N = 20$	$N = 25$	$N = 30$	$N = 35$
carpet	99.00 / 99.45 / 96.77	97.71 / 99.49 / 97.18	99.16 / 99.40 / 97.18	91.93 / 99.20 / 96.00	87.96 / 98.94 / 95.14	98.39 / 99.43 / 97.05	98.84 / 99.51 / 97.20
Grid	100.00 / 98.90 / 94.91	99.92 / 98.91 / 95.61	100.00 / 98.93 / 95.11	99.75 / 98.98 / 95.18	99.83 / 98.96 / 95.50	99.75 / 98.96 / 95.26	99.92 / 98.99 / 95.37
Leather	100.00 / 99.68 / 99.08	100.00 / 99.64 / 98.96	100.00 / 99.67 / 99.10	100.00 / 99.67 / 99.11	100.00 / 99.67 / 99.10	100.00 / 99.66 / 99.05	100.00 / 99.67 / 99.10
Tile	100.00 / 98.69 / 95.13	100.00 / 98.12 / 93.29	100.00 / 98.40 / 94.29	100.00 / 98.80 / 95.74	100.00 / 97.79 / 92.57	100.00 / 98.75 / 95.72	100.00 / 98.78 / 95.21
Wood	99.82 / 97.52 / 91.25	99.47 / 95.84 / 89.61	99.47 / 97.20 / 91.07	99.39 / 94.68 / 87.71	100.00 / 97.01 / 90.99	99.47 / 95.88 / 90.35	99.82 / 96.27 / 89.77
Pill	99.59 / 98.79 / 97.64	99.37 / 98.82 / 97.48	99.73 / 98.79 / 97.58	99.73 / 98.79 / 97.53	99.56 / 98.74 / 97.56	99.65 / 98.79 / 97.66	99.54 / 98.76 / 97.59
Transistor	99.42 / 95.32 / 84.98	99.25 / 95.30 / 84.61	99.58 / 95.18 / 84.63	99.25 / 95.33 / 84.70	99.50 / 95.39 / 85.35	99.42 / 95.27 / 84.57	99.25 / 95.04 / 84.99
Cable	99.03 / 98.62 / 91.10	98.54 / 98.28 / 90.23	98.58 / 98.38 / 90.85	99.06 / 98.47 / 90.97	98.29 / 98.59 / 91.89	98.43 / 98.45 / 91.35	98.76 / 98.43 / 90.89
Zipper	98.35 / 99.25 / 97.56	95.67 / 99.23 / 97.46	98.53 / 99.28 / 97.58	95.98 / 99.27 / 97.54	98.71 / 99.31 / 97.57	95.69 / 99.24 / 97.52	97.11 / 99.24 / 97.39
Toothbrush	99.17 / 99.05 / 93.62	100.00 / 99.21 / 94.84	100.00 / 99.20 / 94.92	100.00 / 99.15 / 94.95	100.00 / 99.21 / 95.06	100.00 / 99.19 / 94.80	100.00 / 99.18 / 94.95
Metal Nut	97.41 / 96.87 / 84.97	97.02 / 96.82 / 85.14	97.90 / 97.01 / 86.09	97.36 / 96.73 / 84.57	97.07 / 97.02 / 84.98	97.17 / 96.78 / 85.32	97.36 / 96.85 / 85.25
Hazelnut	98.57 / 99.10 / 93.95	99.29 / 98.82 / 93.26	98.36 / 98.93 / 93.57	97.89 / 99.16 / 94.10	98.64 / 98.98 / 93.43	98.64 / 98.91 / 93.50	98.68 / 99.01 / 93.98
Screw	96.02 / 99.23 / 96.69	93.85 / 99.28 / 96.91	93.65 / 99.15 / 96.43	94.02 / 99.16 / 96.50	94.16 / 99.16 / 96.52	94.55 / 99.17 / 96.69	93.54 / 99.07 / 96.04
Capsule	94.10 / 96.25 / 88.18	94.34 / 96.84 / 88.30	94.22 / 95.22 / 87.41	93.58 / 96.48 / 88.19	94.18 / 96.76 / 88.65	94.30 / 96.73 / 88.07	94.30 / 96.89 / 88.35
Bottle	99.92 / 98.05 / 92.75	99.84 / 98.06 / 92.74	99.84 / 98.06 / 92.87	99.76 / 98.08 / 92.78	100.00 / 98.06 / 92.72	99.84 / 98.06 / 92.81	99.84 / 98.03 / 92.66
Average	98.69 / 98.32 / 93.24	98.28 / 98.18 / 93.04	98.60 / 98.19 / 93.25	97.85 / 98.13 / 93.04	97.86 / 98.24 / 93.14	98.35 / 98.22 / 93.31	98.46 / 98.25 / 93.25

Table 9. Detailed anomaly detection/localization results with Img-AUROC / Pix-AUROC / PRO (in %) metric on VisA under different N.

Category	$N = 5$	$N = 10$	$N = 15$	$N = 20$	$N = 25$	$N = 30$	$N = 35$
PCB1	97.46 / 99.42 / 93.04	97.51 / 99.49 / 93.29	98.11 / 99.46 / 93.42	97.69 / 99.43 / 93.12	97.70 / 99.44 / 92.76	97.55 / 99.45 / 93.01	98.19 / 99.50 / 93.72
PCB2	99.22 / 98.70 / 87.33	99.11 / 98.27 / 83.12	99.22 / 98.78 / 88.76	98.86 / 98.42 / 84.79	99.43 / 98.46 / 84.68	99.58 / 98.57 / 86.14	99.01 / 98.10 / 82.24
PCB3	98.89 / 99.14 / 95.34	98.85 / 99.10 / 95.41	99.01 / 99.09 / 95.44	99.06 / 99.23 / 95.15	99.17 / 99.18 / 95.30	98.90 / 99.27 / 95.39	98.82 / 99.21 / 95.35
PCB4	99.28 / 94.59 / 80.87	99.21 / 94.96 / 80.08	99.12 / 93.89 / 78.62	99.17 / 94.11 / 79.26	99.25 / 93.78 / 78.00	99.41 / 94.20 / 79.06	99.13 / 94.00 / 79.36
Macaroni 1	96.94 / 99.78 / 98.82	97.04 / 99.76 / 98.64	96.90 / 99.77 / 98.71	97.17 / 99.72 / 98.20	96.63 / 99.78 / 98.66	97.35 / 99.76 / 98.55	96.46 / 99.79 / 98.49
Macaroni 2	90.08 / 99.74 / 98.81	87.98 / 99.73 / 98.90	89.76 / 99.72 / 98.84	89.78 / 99.72 / 98.78	87.63 / 99.72 / 98.82	89.59 / 99.74 / 98.93	89.59 / 99.75 / 98.94
Capsules	95.92 / 99.87 / 98.19	95.77 / 99.86 / 98.58	95.95 / 99.88 / 98.33	95.98 / 99.88 / 98.38	96.88 / 99.87 / 98.44	95.73 / 99.88 / 98.40	96.45 / 99.87 / 98.37
Candle	88.44 / 95.84 / 89.20	87.06 / 95.56 / 89.20	87.40 / 95.68 / 88.93	86.90 / 95.48 / 89.18	86.89 / 95.63 / 88.55	88.06 / 95.58 / 89.18	87.30 / 95.46 / 88.38
Cashew	96.04 / 98.80 / 87.36	95.64 / 98.81 / 88.46	96.70 / 99.17 / 89.77	95.84 / 99.16 / 88.28	95.86 / 99.13 / 88.24	94.82 / 99.05 / 88.19	94.34 / 99.06 / 88.07
Chewinggum	97.28 / 98.29 / 80.10	98.56 / 98.39 / 80.69	98.74 / 98.25 / 80.15	98.74 / 98.32 / 80.41	97.86 / 98.25 / 79.77	98.34 / 98.41 / 80.36	97.70 / 98.50 / 79.96
Fryum	97.84 / 94.71 / 88.93	98.98 / 94.87 / 89.08	99.56 / 94.83 / 89.45	97.92 / 95.23 / 89.71	98.88 / 94.92 / 89.25	99.36 / 94.98 / 89.35	98.74 / 94.81 / 89.26
Pipe fryum	98.68 / 99.41 / 95.82	98.46 / 99.32 / 95.49	98.34 / 99.36 / 95.32	98.92 / 99.27 / 95.93	98.88 / 99.41 / 95.71	98.02 / 99.40 / 95.80	98.34 / 99.31 / 95.37
Average	96.34 / 98.19 / 91.15	96.18 / 98.18 / 90.91	96.57 / 98.16 / 91.31	96.34 / 98.16 / 90.93	96.25 / 98.13 / 90.68	96.39 / 98.19 / 91.03	96.17 / 98.11 / 90.63

Table 10. Detailed anomaly detection/localization results with Img-AUROC / Pix-AUROC / PRO (in %) metric on MvTec-AD under different α.

Category	$\alpha = 1e-2$	$\alpha = 1e-3$	$\alpha = 1e-4$	$\alpha = 1e-5$	$\alpha = 1e-6$	$\alpha = 1e-7$	$\alpha = 1e-8$
carpet	65.17 / 89.93 / 84.99	65.77 / 92.34 / 84.61	79.21 / 96.16 / 86.39	95.95 / 98.68 / 93.79	99.00 / 99.45 / 96.77	97.15 / 99.14 / 96.08	97.99 / 99.28 / 96.94
Grid	100.00 / 99.31 / 93.59	100.00 / 99.31 / 93.39	98.50 / 96.16 / 84.33	99.83 / 98.03 / 89.63	100.00 / 98.90 / 94.91	100.00 / 99.21 / 96.25	99.42 / 99.20 / 95.41
Leather	100.00 / 99.67 / 98.12	100.00 / 99.68 / 98.52	100.00 / 99.69 / 98.19	100.00 / 99.59 / 99.15	100.00 / 99.68 / 99.08	100.00 / 99.52 / 97.91	100.00 / 99.62 / 98.83
Tile	96.07 / 97.76 / 92.53	98.70 / 97.83 / 92.82	94.23 / 87.81 / 72.86	100.00 / 97.82 / 94.01	100.00 / 98.69 / 95.13	99.13 / 97.68 / 93.36	99.42 / 96.86 / 91.80
Wood	100.00 / 95.51 / 88.55	96.75 / 93.52 / 86.31	97.46 / 91.57 / 82.31	98.77 / 95.42 / 90.09	99.82 / 97.52 / 91.25	100.00 / 95.61 / 89.36	100.00 / 94.72 / 86.83
Pill	94.76 / 90.26 / 81.69	94.33 / 88.49 / 79.56	98.94 / 98.68 / 96.56	98.15 / 98.59 / 97.54	99.59 / 98.79 / 97.64	99.07 / 97.69 / 92.43	97.27 / 89.58 / 83.11
Transistor	99.75 / 96.61 / 83.25	99.71 / 96.91 / 83.75	98.88 / 96.05 / 85.53	99.12 / 95.10 / 85.02	99.42 / 95.32 / 84.98	99.87 / 96.36 / 84.57	99.87 / 95.43 / 83.58
Cable	98.16 / 98.39 / 89.14	98.20 / 98.43 / 88.79	99.53 / 98.07 / 89.13	95.91 / 98.02 / 90.50	99.03 / 98.62 / 91.10	98.84 / 98.24 / 86.98	98.63 / 98.18 / 85.29
Zipper	98.69 / 99.32 / 94.01	96.32 / 99.31 / 93.77	95.04 / 98.60 / 95.96	98.48 / 98.86 / 96.81	98.35 / 99.25 / 97.56	97.11 / 99.31 / 97.30	97.64 / 99.33 / 97.19
Toothbrush	100.00 / 99.22 / 95.66	99.44 / 99.33 / 95.64	99.72 / 99.26 / 95.32	100.00 / 99.17 / 95.01	99.17 / 99.05 / 93.62	100.00 / 99.18 / 94.67	100.00 / 99.18 / 94.49
Metal Nut	95.65 / 96.42 / 74.36	95.89 / 96.40 / 75.66	95.99 / 94.44 / 83.17	93.26 / 94.62 / 83.35	97.41 / 96.87 / 84.97	96.14 / 96.38 / 75.24	93.21 / 95.26 / 69.19
Hazelnut	97.54 / 98.41 / 88.22	97.43 / 98.31 / 82.92	97.64 / 98.19 / 88.04	98.82 / 99.26 / 94.23	98.57 / 99.10 / 93.95	95.64 / 95.54 / 86.40	97.86 / 92.97 / 81.64
Screw	91.10 / 98.98 / 95.57	94.45 / 98.95 / 95.42	93.20 / 98.97 / 95.58	96.23 / 99.53 / 97.80	96.02 / 99.23 / 96.69	86.94 / 98.57 / 93.09	77.06 / 97.79 / 89.28
Capsule	95.69 / 98.12 / 90.58	94.81 / 98.18 / 90.62	94.58 / 96.84 / 88.57	94.26 / 96.41 / 87.64	94.10 / 96.25 / 88.18	96.89 / 98.61 / 93.25	97.05 / 98.56 / 93.39
Bottle	99.92 / 98.68 / 88.78	100.00 / 98.69 / 89.27	100.00 / 98.22 / 93.22	100.00 / 97.64 / 91.75	99.92 / 98.05 / 92.75	100.00 / 98.66 / 93.48	100.00 / 98.63 / 92.82
Average	95.50 / 97.10 / 89.27	95.45 / 97.04 / 88.74	96.19 / 96.58 / 89.01	97.92 / 97.78 / 92.42	98.69 / 98.32 / 93.24	97.79 / 97.98 / 91.36	97.03 / 96.97 / 89.32

E. More Qualitative Results

This section reports more qualitative Results on the MvTec-AD and VisA dataset. Figure 6 depicted results from different categories of the compared methods. Figure 7 depicted intermediate step and anomaly localization results under different sampling function and different μ_s. Due to the different sparsity of different channel masks, some $\hat{X}$ will show color cast or brightness reduction.
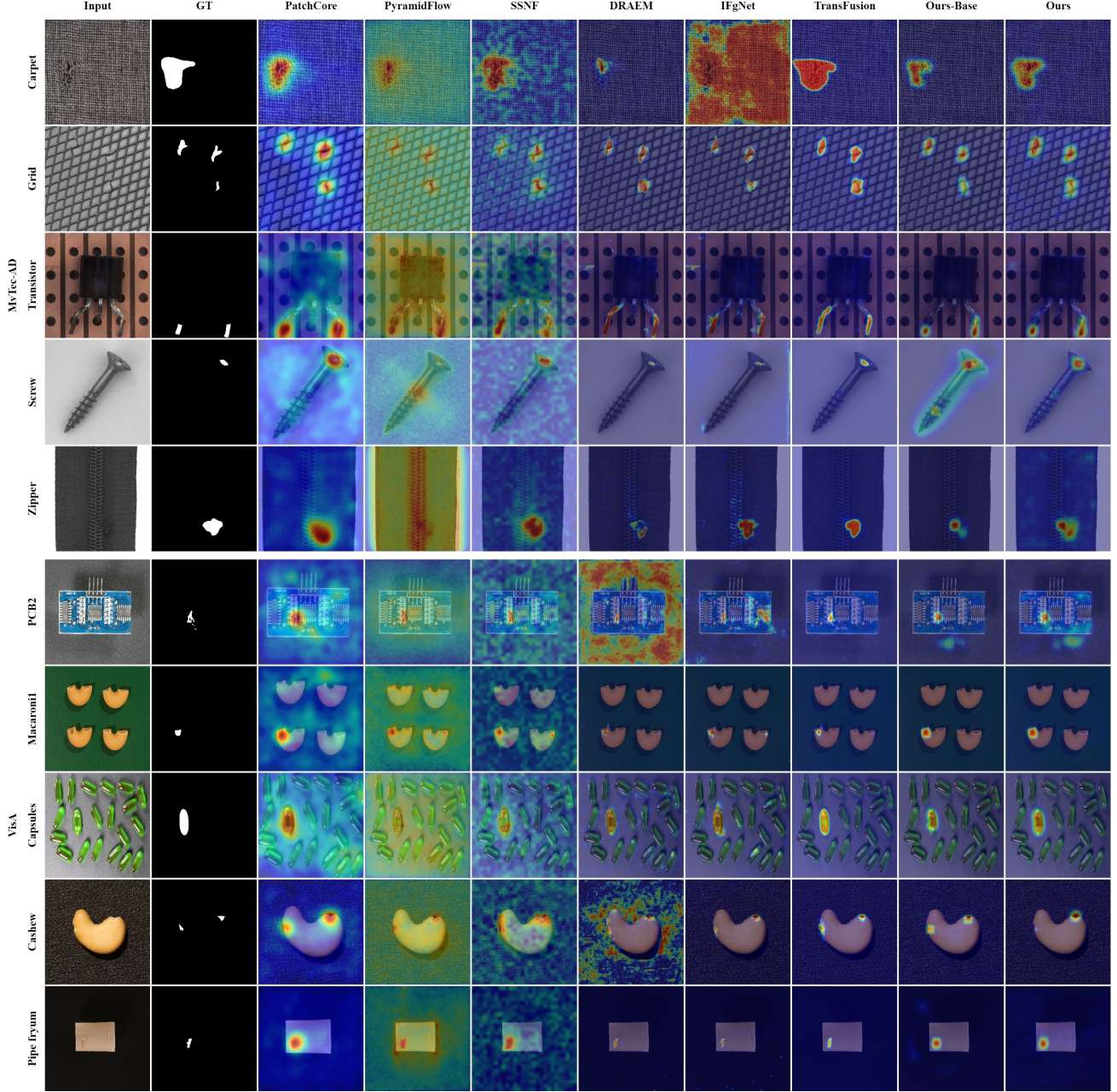

Figure 6. Qualitative comparison for anomaly localization on two datasets. From left to right: the input images, ground-truth, and the anomaly maps produced by all compared methods.
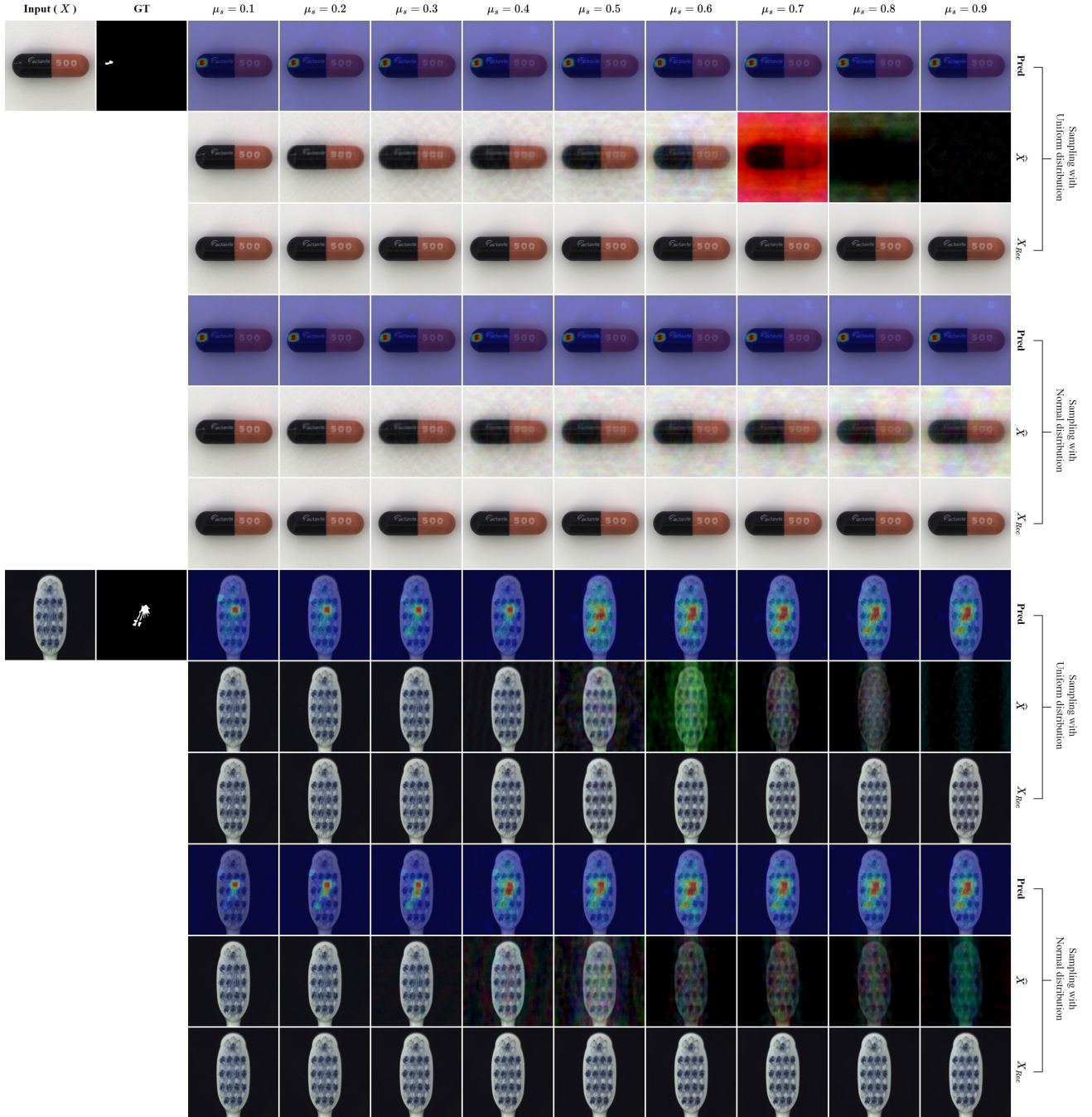
Figure 7. Intermediate step and anomaly localization results under different sampling function and different μ_s.