Beyond Normal References: Discriminative Few-Shot Anomaly Detection

Huan Wang[1,2]*, **Jun Shen**[1]†, **Jun Yan**[1], **Guansong Pang**[2]†
[1]University of Wollongong, Australia
[2]Singapore Management University, Singapore

https://github.com/mala-lab/IDEAL

Abstract

This paper considers a practical few-shot anomaly detection (FSAD) setting, termed **discriminative FSAD**, where a limited number of both normal and anomalous examples are available as references during inference. Existing FSAD methods rely on normal-only references through normality matching, ignoring the discriminative clues in anomalous references, while directly fitting both references can overfit to the seen anomalies. We introduce **IDEAL**, an **i**ntrinsic **de**vi**a**tion **l**earning framework that leverages both reference types to learn intrinsic deviation patterns characterizing generalizable abnormality as deviations from normality. IDEAL decomposes the learning process into two novel components: 1) a **Normal Variation Eraser** to suppress nuisance normal variations that may lead to noisy deviations from normality, thereby highlighting anomaly-relevant deviation representations; 2) an **Intrinsic Deviation Encoder** to decompose these denoised deviation representations into intrinsic deviation vectors capturing the most discriminative orthogonal deviation directions. During inference, IDEAL scores query-to-normal deviations preserved after projection onto the learned intrinsic deviation vectors, enabling generalization for both seen and unseen anomalies. Extensive experiments on eight real-world datasets show that IDEAL generalizes effectively to unseen anomalies and consistently outperforms existing state-of-the-art FSAD methods.

1 Introduction

Anomaly detection (AD) [11, 54] aims to identify data points that significantly depart from expected patterns of normal data and has achieved a bright promise in many applications, such as industrial inspection [37, 72] and medical image diagnosis [23, 30, 85]. Conventional AD methods [14, 16, 26, 57, 71, 78, 80, 82], which adopt a full-shot setting, rely on access to a large collection of normal training samples. To alleviate this requirement, Few-Shot Anomaly Detection (FSAD), which assumes the availability of only a limited number of training samples, has emerged as a popular direction [29, 32, 36, 44, 45, 74, 77, 84, 93]. However, most existing FSAD methods still follow a specialist paradigm, where a dataset-specific training is required for each target data using few-shot normal samples [22, 29, 36, 75]. Although effective, this one-model-per-dataset paradigm requires substantial re-engineering and repeated re-training for each new deployment. This motivates the development of generalist FSAD [77, 84, 93], which aims to learn a single detector capable of generalizing across diverse datasets without specific retraining/tuning. It achieves this by leveraging few-shot normal examples from the target data as in-context normal references at test time.

While these normal references provide a profile of normality, they do not offer explicit discriminative evidence for abnormality [18, 55, 79]. In practice, however, a limited number of anomalous examples can often be made available from historical records (*e.g.*, images of confirmed tumor cases or defective products flagged by human inspectors). These anomalous examples encode valuable prior knowledge

*This work was done while Huan was visiting at SMU (hw226@uowmail.edu.au).
†Corresponding Authors: Guansong Pang (gspang@smu.edu.sg) and Jun Shen (jshen@uow.edu.au).

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

about anomalies in the target data, yet existing FSAD methods—designed to rely solely on normal references—cannot leverage this information.

Motivated by this gap, we consider a practical FSAD setting, termed **discriminative FSAD**, where a limited number of both normal and anomalous examples are available as references at inference time. Despite their potential utility, the use of these few-shot anomalous samples as additional references introduces a fundamental challenge: the few-shot abnormal references are inherently sparse and often fail to illustrate the full spectrum of anomaly types due to the strong heterogeneity in their distributions [54]. Consequently, naively conditioning on such references can bias the detector toward seen anomaly patterns, leading to overfitting and degraded generalization to unseen anomalies. As illustrated in Fig. 1 (**b**) and (**c**), we implement KNN similarity matching to examine the feasibility of directly using the abnormal references in FSAD. Direct matching of the input image to either anomalous-only references or both normal-anomalous references can produce noisy and spurious anomaly activations, due to some major differences of abnormality in the input image and anomalous reference image. This indicates that abnormal references should not be treated as matching templates to identify anomalies, because they cover only some specific anomalous types.

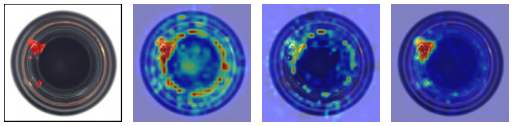

Figure 1: Anomaly score maps of **(a)** an input image using direct similarity matching of **(b)** anomalous-only references and **(c)** both anomalous and normal references, compared to **(d)** our method IDEAL under a 1-shot normal and anomalous reference setting. Two simple reference matching methods produce noisy and spurious anomaly responses, whereas IDEAL yields substantially cleaner activations. Note that, to illustrate the anomaly heterogeneity issue, the anomalous reference was intentionally chosen to be different from the abnormality in this input.

To address this issue, we propose **IDEAL**, an **i**ntrinsic **de**vi**a**tion **l**earning framework for discriminative FSAD, which leverages both normal and abnormal references to learn intrinsic deviation patterns that characterize generalizable abnormality as deviations from normality, thereby supporting the detection of both seen and unseen anomalies. The key insight of IDEAL is to learn a set of *intrinsic deviation patterns* that capture discriminative, relative deviations from normality shared by both seen and unseen anomalies, even when unseen anomalies do not resemble the seen anomalies. Specifically, IDEAL decomposes the learning process into two novel components: 1) a **Normal Variation Eraser (NVE)**, which suppresses nuisance normal variations (*e.g.*, illumination/texture shifts) that may lead to noisy deviations from normality, thereby highlighting anomaly-relevant deviation representations; and 2) an **Intrinsic Deviation Encoder (IDE)**, which learns to decompose these denoised deviations into intrinsic deviation vectors capturing the most discriminative orthogonal deviation directions. This allows IDEAL to identify anomalies by measuring how well each query-to-normal deviation is preserved after projection onto these learned intrinsic deviation vectors, rather than by matching the query to the limited references, enabling effective generalization to unseen anomalies, as shown in Fig. 1 (**d**).

The main contributions of this paper are summarized as follows:

- We formulate discriminative FSAD, a practical extension of conventional FSAD, where limited normal and anomalous samples are both available as in-context references, aiming to exploit the discriminative clues of abnormal references relative to normal references.
- We propose an intrinsic deviation learning (IDEAL) framework for discriminative FSAD, which leverages both reference types to learn intrinsic deviation patterns characterizing generalizable abnormality as deviations from normality. By measuring query-to-normal deviations preserved after projection onto these learned intrinsic deviation vectors, IDEAL effectively detects both seen and unseen anomalies.
- Comprehensive experiments on eight real-world datasets show the superiority of IDEAL in terms of performance, efficiency, and generalizability against state-of-the-art FSAD methods. Extensive ablation studies also validate the contribution of the two proposed components.

2 Related Work

Conventional Anomaly Detection. Anomaly detection approaches have been developed for diverse data and application scenarios [11, 12, 54, 60]. In particular, one-class classification methods aim to

learn a compact normality boundary from the given category using support vectors [7, 13, 58, 81]. Reconstruction-based methods [2, 15, 25, 27, 40, 43, 56, 61] capture normal patterns by reconstructing normal samples in feature space, with reconstruction discrepancies serving as indicators of abnormality. Distance-based methods detect anomalies by measuring the feature-level discrepancy between the query and the normal training distribution [14, 16, 53, 57, 65]. Knowledge distillation methods [9, 10, 17, 59, 87] train a student network on normal samples to mimic the feature responses of a pre-trained teacher network, and detect anomalies using the resulting teacher-student discrepancies. Synthetic-based methods [35, 41, 83, 86] introduce anomaly-like supervision by corrupting normal samples or perturbing normal features, and then train the model to distinguish normal from synthetic abnormal patterns. Further, there have been supervised methods [18, 55, 58, 70, 91] that explored the use of few-shot labeled anomalies to improve anomaly detection. However, these methods require large normal training data and dataset-specific training for each deployment, limiting their applicability in real-world scenarios where target data are difficult/costly to access [3, 49].

Few-Shot Anomaly Detection. Few-shot anomaly detection (FSAD) aims to reduce the dependence on full-shot normal training data by using a limited number of training samples. Existing FSAD methods are mainly divided into two aspects: specialist FSAD and generalist FSAD. Specialist FSAD [22, 29, 36, 38, 64, 66, 74, 75] focuses on adapting a detector for each category/dataset using few available normal samples. For example, RegAD [29] learns normal representations via a registration proxy task and detects anomalies by comparing registered query-support features. However, these methods still rely on category- or dataset-specific fitting, leading to a specialist one-model-per-dataset paradigm with limited scalability. To overcome this limitation, generalist FSAD [19, 32, 44, 45, 77, 84, 93] aims to learn a single detector that can generalize across diverse datasets without target-specific retraining. InCTRL [93] performs in-context residual learning with few-shot normal sample prompts and ResAD [77] applies residual features to eliminate domain dependencies, to achieve generalizable detection. From the perspective of using references, NAGL [73] extends FSAD by introducing anomalous references and performing residual mining and learning to detect anomalies. In addition, many zero-shot AD methods [24, 32, 34, 46, 76, 90, 92] have been introduced recently, sharing the spirit of adapting large pre-trained models to detect anomalies.

3 Problem Formulation

Generalist FSAD. According to [93], given a set of training data $\mathcal{D}_{tr} = \{\mathcal{X}, \mathcal{Y}\}$, where $\mathcal{X}$ consists of several normal and abnormal images and $\mathcal{Y}$ denotes corresponding labels. At inference, x is a test query from a target class, accompanied by few-shot normal references $\mathcal{S}^n$. The goal is to train a detection model on $\mathcal{D}_{tr}$ that can generalize to detect anomalies of x with $\mathcal{S}^n$.

Discriminative FSAD. Given a set of auxiliary training data $\mathcal{D}_{tr} = \{\mathcal{X}, \mathcal{Y}\}$, where $\mathcal{X} = \{\mathbf{x}_z\}_{z=1}^{Z}$ consists of Z normal and abnormal samples, $\mathcal{Y} = \{\mathbf{y}_z\}_{z=1}^{Z}$ as the corresponding labels ($\mathbf{y}_z = 0$ indicates normal, $\mathbf{y}_z = 1$ signifies abnormal), during **training**, we first randomly sample a few-shot reference set $\mathcal{S} = \{\mathcal{S}^n, \mathcal{S}^a\}$ from $\mathcal{D}_{tr}$ for query input $\mathbf{x}^q \in \mathcal{X}$, where $\mathcal{S}^n = \{\mathbf{s}_l^n\}_{l=1}^{L_1}$ and $\mathcal{S}^a = \{\mathbf{s}_l^a\}_{l=1}^{L_2}$ represent L_1-shot normal and L_2-shot abnormal references, respectively; we then train a detection model $\Phi : \mathcal{D}_{tr} \to \mathbb{R}$ based on different input **episodes** $\mathbf{E} = \{\mathbf{x}^q, \mathcal{S}^n, \mathcal{S}^a\}$. At **inference**, the model Φ is directly evaluated on the target dataset without any further tuning/retraining. *Note that to replicate real-world scenarios where only a few reference examples are available and the same ones are used across all test samples, for any new dataset c, the few-shot reference set $\mathcal{S}_c = \{\mathcal{S}_c^n, \mathcal{S}_c^a\}$ is fixed, where $\mathcal{S}_c^a$ is drawn only once from one randomly selected anomaly type and $\mathcal{S}_c^n$ consists of a few normal references sampled from the normal class.* This setup differs from [73], which dynamically resamples the abnormal reference examples for each test image based on its anomaly type, thereby leaking information about test-time anomaly type and also increasing the reference set size. We set $L_1 > L_2$, motivated by the fact that anomalies are rare in real-world applications.

Further, for each episode $\mathbf{E} = \{\mathbf{x}^q, \mathcal{S}^n, \mathcal{S}^a\}$, including a query image $\mathbf{x}^q \in \mathbb{R}^{H \times W \times 3}$ with a label $\mathbf{y}^q$ and a ground-truth mask $\mathcal{M}^q$ (each normal image is equipped with a zero mask), L_1-shot normal references $\mathcal{S}^n$ and L_2-shot abnormal references $\mathcal{S}^a$. We use a pre-trained encoder $\mathcal{E}(\cdot)$ to extract query features $\mathcal{F}^q = \mathcal{E}(\mathbf{x}^q) = \{f_i^q\}_{i=1}^{N}$, normal features $\mathcal{F}^n = \mathcal{E}(\mathcal{S}^n) = \{f_i^n\}_{i=1}^{L_1 \cdot N}$, and abnormal features $\mathcal{F}^a = \mathcal{E}(\mathcal{S}^a) = \{f_i^a\}_{i=1}^{L_2 \cdot N}$, where $f_i^q, f_i^n, f_i^a \in \mathbb{R}^{\mathbb{C}}$, N is the number of patches, $\mathbb{C}$ is the dimension. For $\mathcal{S}^a$, we down-sample its anomaly masks to match $\mathcal{F}^a$, yielding $\mathcal{M}^a \in \{0, 1\}^{L_2 \cdot N}$.

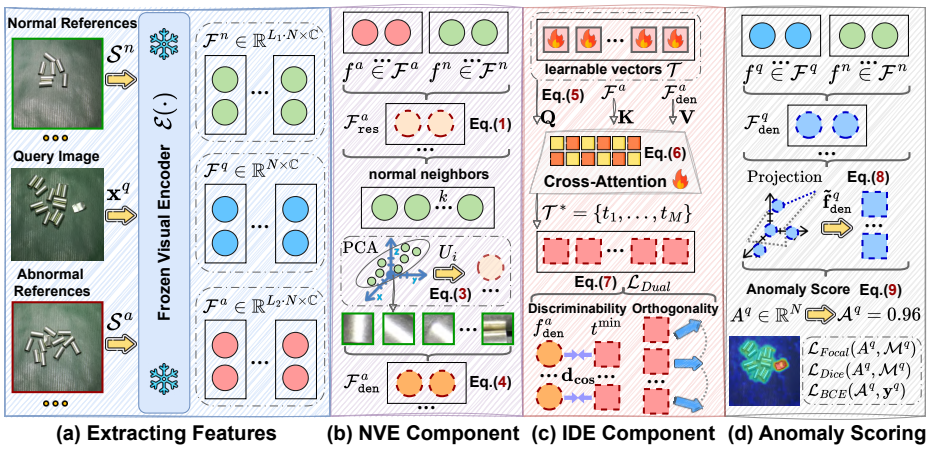

Figure 2: Overview of IDEAL (Algorithm 1). **(a)** Given each input $\mathbf{E} = \{\mathbf{x}^q, \mathcal{S}^n, \mathcal{S}^a\}$, we employ a pre-trained encoder $\mathcal{E}(\cdot)$ to extract features. We then introduce **(b)** a Normal Variation Eraser (Sec. 4.1) and **(c)** an Intrinsic Deviation Encoder (Sec. 4.2) to learn intrinsic deviation vectors from both references. Finally, **(d)** anomalies are detected by measuring query-to-normal deviations preserved after projection onto these learned intrinsic deviation vectors (Sec. 4.3).

4 IDEAL: Intrinsic Deviation Learning for Discriminative FSAD

Overview of Our Proposed IDEAL Framework. To effectively leverage the discriminative clues of abnormal references relative to normal references, IDEAL learns intrinsic deviation patterns characterizing generalizable abnormality as deviations from normality, to detect both seen and unseen anomalies. As shown in Fig. 2, given a query sample accompanied by few-shot normal and abnormal references, IDEAL first extracts patch-level features via a pre-trained visual encoder. Then, IDEAL proceeds with two key components: 1) NVE (Sec. 4.1), which suppresses nuisance normal variations that may lead to noisy deviations from normality to highlight anomaly-relevant deviation representations; 2) IDE (Sec. 4.2), which learns to decompose these denoised deviation representations into intrinsic deviation vectors capturing the most discriminative orthogonal deviation directions. Finally, the anomaly score is obtained by measuring query-to-normal deviations preserved after projection onto the learned intrinsic deviation vectors (Sec. 4.3).

4.1 Normal Variation Eraser (NVE)

The key intuition of NVE is to isolate anomaly-relevant deviation representations from irrelevant/noisy deviations by eliminating nuisance normal variations that may lead to these noisy deviations. Hence, NVE operates in two steps: 1) it extracts feature residual-based deviations by contrasting abnormal and normal references, canceling out class-specific content; 2) it suppresses nuisance normal variations embedded in the residual deviations, producing denoised deviations that more faithfully reflect anomalous departures from normality.

Residual Deviations. For each patch-level abnormal feature f_i^a of $\mathcal{F}^a$, we can search the nearest normal reference feature $f_{i,\min}^n := \texttt{argmin}_{f^n \in \mathcal{F}^n} \mathbf{d_{cos}}(f^n, f_i^a)$, where $\mathbf{d_{cos}}(\cdot,\cdot)$ denotes the cosine function and $i \in \{1, 2, \ldots, L_2 \cdot N\}$. We define residual deviations between $\mathcal{F}^a$ and $\mathcal{F}^n$ as:

$$\mathcal{F}_{\texttt{res}}^a = \{f_i^a - \texttt{argmin}_{f^n \in \mathcal{F}^n} \mathbf{d_{cos}}(f^n, f_i^a)\}_{i=1}^{L_2 \cdot N} \in \mathbb{R}^{L_2 \cdot N \times \mathbb{C}}. \tag{1}$$

Denoised Deviations. Although the residual deviations $\mathcal{F}_{\texttt{res}}^a$ reveal departures from normality, they may still contain nuisance normal variations that may lead to noisy deviations from normality. To estimate those normal variation around each abnormal feature f_i^a of $\mathcal{F}^a$, we first retrieve its top-k nearest normal neighbors $\mathcal{F}_{i,k}^n$ from the normal features $\mathcal{F}^n$ as follows:

$$\mathcal{F}_{i,k}^n = \texttt{argsort}_{f^n \in \mathcal{F}^n}(\mathbf{d_{cos}}(f^n, f_i^a), k) = \{f_{i,j}^n\}_{j=1}^k \in \mathbb{R}^{k \times \mathbb{C}}. \tag{2}$$

Based on $\mathcal{F}_{i,k}^n$, we can calculate the empirical mean $\mu_i \in \mathbb{R}^{\mathbb{C}}$ and covariance matrix $\Sigma_i \in \mathbb{R}^{\mathbb{C} \times \mathbb{C}}$. We then estimate a local normal variation subspace $\mathcal{R}_i$ from the centered neighbors $\{f_{i,j}^n - \mu_i\}_{j=1}^k$

using Principal Component Analysis (PCA) [1, 67], which yields an orthonormal basis U_i for $\mathcal{R}_i$. Accordingly, each local normal neighbor $f^n_{i,j}$ is approximated as:

$$f^n_{i,j} = \mu_i + U_i v_{i,j} + \epsilon_{i,j}, \quad \text{s.t. } v_{i,j} \sim \mathcal{N}(0, \mathbf{I}_r),\ \epsilon_{i,j} \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_{\mathbb{C}}), \tag{3}$$

where $U_i \in \mathbb{R}^{\mathbb{C}\times r}$ contains top-r eigenvectors of Σ_i, $v_{i,j} \in \mathbb{R}^r$ is a latent variable, $\epsilon_{i,j}$ is an isotropic noise term, $\mathbf{I}_r$ is a $r \times r$ identity matrix, and $\mathbf{I}_{\mathbb{C}}$ is a $\mathbb{C} \times \mathbb{C}$ identity matrix. The columns of $U_i \in \mathbb{R}^{\mathbb{C}\times r}$ form an orthonormal basis for the local normal variation subspace $\mathcal{R}_i$, where r denotes the number of retained principal components in $\mathcal{R}_i$ and typically satisfies $r < \min(k-1, \mathbb{C})$. Further, for each residual deviation $f^a_{i,\texttt{res}} = (f^a_i - f^n_{i,\min}) \in \mathcal{F}^a_{\texttt{res}}$ in Eq. (1), we suppress its noisy component aligned with the local normal variation subspace $\mathcal{R}_i$, obtaining the denoised deviations:

$$\mathcal{F}^a_{\texttt{den}} = \{f^a_{i,\texttt{res}} - \alpha U_i U_i^\top f^a_{i,\texttt{res}}\}_{i=1}^{L_2\cdot N} = \{f^a_{i,\texttt{den}}\}_{i=1}^{L_2\cdot N} \in \mathbb{R}^{L_2\cdot N\times \mathbb{C}}, \tag{4}$$

where $\alpha \in (0, 1]$ is used to controls the degree of elimination. In this way, the denoised deviations $\mathcal{F}^a_{\texttt{den}} \in \mathbb{R}^{L_2\cdot N\times\mathbb{C}}$ suppress noisy components aligned with the local normal variation subspace while preserving anomaly-relevant components that are less explainable by normal variations.

4.2 Intrinsic Deviation Encoder (IDE)

While denoised deviations highlight the anomalous departures from normality, directly employing similarity matching to compare query-to-normal deviations with these denoised deviations remains fragile, as it still depends on the limited abnormal references. To address this limitation, we introduce an Intrinsic Deviation Encoder (IDE) to learn intrinsic deviation vectors decomposed from these denoised deviation representations across different input episodes, thereby capturing the most discriminative orthogonal deviation directions shared by different anomalies. Specifically, we employ a cross-attention [69] mechanism to capture such intrinsic deviation vectors, where the three matrices $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ in the attention are defined as:

$$\mathbf{Q} = \omega^Q(\mathcal{T}) \in \mathbb{R}^{M\times\mathbb{C}},\ \mathbf{K} = \omega^K(\mathcal{F}^a) \in \mathbb{R}^{L_2\cdot N\times\mathbb{C}},\ \mathbf{V} = \omega^V(\mathcal{F}^a_{\texttt{den}}) \in \mathbb{R}^{L_2\cdot N\times\mathbb{C}}, \tag{5}$$

where $\mathcal{T} \in \mathbb{R}^{M\times\mathbb{C}}$ consists of M learnable vectors, $\{\omega^Q, \omega^K, \omega^V\}$ transform the input features into query, key, and value vectors. Then, IDE learns intrinsic deviation vectors $\mathcal{T}^*$ by:

$$\mathcal{T}^* = \{t_m\}_{1\le m\le M} = \texttt{FFN}(\texttt{Softmax}(\mathbf{Q}\mathbf{K}^\top/\sqrt{\mathbb{C}} + \gamma(1-\mathcal{M}^a))\mathbf{V}), \tag{6}$$

where $\mathcal{M}^a \in \{0,1\}^{L_2\cdot N}$ is the patch-level binary anomaly masks aligned with $\mathcal{F}^a$, $\gamma \to -\infty$ is used to ensure that the attention weights of normal patches in $\mathcal{F}^a$ are close to zero, and $\texttt{FFN}(\cdot)$ represents a feed forward network. $\mathcal{T}^* \in \mathbb{R}^{M\times\mathbb{C}}$ represents M intrinsic deviation vectors of abnormality relative to normality. For brevity, we omit the residual-path operations in Eq. (6).

On the one hand, we expect to narrow the distance between each individual element in $\mathcal{F}^a_{\texttt{den}}$ and the corresponding nearest $t^{\min}$ in $\mathcal{T}^*$ to enhance anomaly-aware discriminative ability. On the other hand, we emphasize mutual orthogonality among these intrinsic deviation vectors $\mathcal{T}^*$ to reduce redundancy. Therefore, we encourage the learning of $\mathcal{T}^*$ by:

$$\mathcal{L}_{Dual} = \lambda_1 \underbrace{\frac{1}{\sum_{i=1}^{L_2\cdot N}\mathcal{M}^a_i}\sum_{i=1}^{L_2\cdot N}\mathcal{M}^a_i \mathbf{d}_{\cos}(f^a_{i,\texttt{den}}, t^{\min}_i)}_{\text{Discriminability}} + \lambda_2 \underbrace{\frac{1}{M(M-1)}\sum_{m_1\neq m_2}\left(\frac{t_{m_1}\cdot t_{m_2}}{\|t_{m_1}\|\times\|t_{m_2}\|}\right)^2}_{\text{Orthogonality}}, \tag{7}$$

where $\mathcal{M}^a_i \in \{0,1\}$ is the corresponding mask value in $\mathcal{M}^a$, $t^{\min}_i := \texttt{argmin}_{t_m\in\mathcal{T}^*}\mathbf{d}_{\cos}(t_m, f^a_{i,\texttt{den}})$ represents the nearest deviation pattern to the $f^a_{i,\texttt{den}}$, λ_1 and λ_2 are hyper-parameters used to balance these loss terms. $\mathcal{L}_{Dual}$ promotes the learning of the intrinsic deviation vectors $\mathcal{T}^*$ that capture most discriminative orthogonal deviation directions from normality while maintaining mutual diversity, thereby improving the generalizable ability for detecting diverse abnormal patterns.

4.3 Anomaly Scoring with Projection

For anomaly scoring, we first extract the denoised deviations for each query $\mathbf{x}^q$: $\mathcal{F}^q_{\texttt{den}} = \{f^q_{i,\texttt{den}}\}_{i=1}^N \in \mathbb{R}^{N\times\mathbb{C}}$. Next, we project these denoised deviations $f^q_{i,\texttt{den}}$ onto the learned intrinsic deviation vectors:

$$\tilde{\mathbf{f}}^q_{i,\texttt{den}} = \texttt{Proj}(f^q_{i,\texttt{den}}) := \sum_{m=1}^{M}(\langle f^q_{i,\texttt{den}}, t_m\rangle/\langle t_m, t_m\rangle)t_m, \tag{8}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product operation. This projection retains only the relevant attributes of anomalous deviations, as irrelevant attributes are nearly orthogonal to the learned intrinsic deviation vectors. Finally, the anomaly score for each patch in the query $\mathbf{x}^q$ is treated as a complementary score of normal matching and deviation alignment, which can be calculated by:

$$A_i^q = \frac{1}{2}(1 - \mathbf{d_{cos}}(f_{i,\mathtt{den}}^q, \tilde{\mathbf{f}}_{i,\mathtt{den}}^q) + \mathbf{d_{cos}}(f_i^q, f_{i,\min}^n)) \;\Rightarrow\; A^q = \{A_i^q\}_{i=1}^N \in \mathbb{R}^N, \tag{9}$$

where $A_i^q = 0$ indicates the most normal region and $A_i^q = 1$ indicates the most anomalous region, and $f_{i,\min}^n$ is the nearest normal reference feature to f_i^q. Following [31], we obtain the final image-level anomaly score $\mathcal{A}^q$ by averaging the top 1% highest anomaly patch scores in the A^q.

4.4 Training and Inference

Training on Auxiliary Datasets. IDEAL is trained on auxiliary data $\mathcal{D}_{tr}$ in an episodic manner. For each training episode $\mathbf{E} = \{\mathbf{x}^q, \mathcal{S}^n, \mathcal{S}^a\}$ from $\mathcal{D}_{tr}$, we optimize IDEAL with two objectives: 1) enforcing consistency between the predicted score map A^q in Eq. (9) and the query's ground-truth mask $\mathcal{M}^q$; and 2) supervising the image-level anomaly score $\mathcal{A}^q$ with the query's label $\mathbf{y}^q$. Therefore, we employ Focal loss [39] and Dice loss [51] to optimize the segmentation branch, while using BCE (Binary Cross-Entropy) loss [47] to supervise the classification branch, formulated as:

$$\mathcal{L}_{train} = \mathcal{L}_{Focal}(A^q, \mathcal{M}^q) + \mathcal{L}_{Dice}(A^q, \mathcal{M}^q) + \mathcal{L}_{BCE}(\mathcal{A}^q, \mathbf{y}^q) + \mathcal{L}_{Dual}. \tag{10}$$

Inference. During inference, for any new dataset c, we assume a fixed reference set $\mathcal{S}_c = \{\mathcal{S}_c^n, \mathcal{S}_c^a\}$, where $\mathcal{S}_c^a$ is drawn from one randomly selected anomaly type and $\mathcal{S}_c^n$ consists of a few normal images randomly sampled from the normal class. The resulting reference set $\mathcal{S}_c$ is then fixed and used in the anomaly scoring of all test samples. Given a test query x, IDEAL yields a deviation-guided anomaly score map A^x based on the input $\{x, \mathcal{S}_c^n, \mathcal{S}_c^a\}$. Finally, A^x is upsampled to the original image resolution $\mathbb{R}^{H \times W}$ for anomaly localization.

4.5 Generalization Bound

We provide insights into the source-target generalization analysis of IDEAL. A detailed description and derivations of this analysis can be found in Appendix B.

Theorem 4.1 (Generalization Bound of IDEAL)**.** *Consider there are K labeled source domains. Let $\mathbb{P}_{S_1}, \ldots, \mathbb{P}_{S_K}$ be the corresponding source episode distributions, and $\mathbb{P}_T$ be the target episode distribution. Denote the model Φ as the hypothesis from $\mathcal{H}$ and d be the VC-dimension of $\mathcal{H}$. The total number of episodes over all K source domains is n. Then, for any source weighting vector $\beta = (\beta_1, \ldots, \beta_K) \in \Delta_K$, with probability at least $1 - \delta$, the following holds for any $\Phi \in \mathcal{H}$,*

$$\epsilon_T(\Phi) \leq \hat{\epsilon}_\beta(\Phi) + \sqrt{\frac{1}{2}\left(\sum_{k=1}^K \frac{\beta_k^2}{n_k}\right) \log \frac{M}{\delta}} + 2\sum_{k=1}^K \beta_k \sqrt{\frac{d}{n_k} \log \frac{en_k}{d}} + \mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta) + \Omega_\beta. \tag{11}$$

Theorem 4.1 indicates that the generalization bound of IDEAL is associated with the empirical source risk $\hat{\epsilon}_\beta(\Phi)$, the number M of intrinsic deviation vectors, the source-target data discrepancy $\mathrm{D}_{\mathcal{H},\ell}$, and the joint approximation term Ω_β. Therefore, a tighter bound can be expected when Φ attains low empirical source risk and the number M of intrinsic deviation vectors is properly chosen.

5 Experiments

5.1 Experimental Setup

Datasets. To verify the effectiveness of IDEAL, we conduct experiments across eight AD datasets, including five industrial datasets: MVTecAD [8], VisA [94], AITEX [62], MPDD [33], BTAD [52]; and three medical image datasets: BraTS [50] (for brain tumor segmentation), Liver [4] (for liver tumor segmentation), RESC [28] (for retinal edema segmentation). To assess the AD performance, we adopt a cross-dataset evaluation: the model is trained on MVTecAD as the source dataset and then directly evaluated on the test sets of the other seven datasets without any further training, following prior work [77, 93]. When evaluating on MVTecAD, the model is trained on VisA. During inference, unlike [73], which samples different abnormal references based on the query's anomaly type, we sample them once from each test dataset and keep them fixed for all test samples in the same dataset.

Table 1: **Comparison results (*General*) with AUROC metric** under various few-shot AD settings, where $(\cdot/\cdot)$ means image-level and pixel-level AUROCs, `N`i and `A`i denote the i-shot of normal and abnormal references. Best results are in **red bold**, with second best results in **blue bold**.

SETUPS	METHODS	MVTECAD	VISA	AITEX	MPDD	BTAD	BRATS
N1	RegAD [29]	(91.6 / 93.1)	(83.3 / 91.6)	(72.0 / 80.2)	(69.7 / 92.5)	(86.6 / 94.5)	(64.2 / 92.2)
	PromptAD [36]	(92.8 / 93.4)	(83.8 / 89.5)	(71.9 / **81.4**)	(75.8 / 95.1)	(89.6 / 92.4)	(68.7 / 91.9)
	WinCLIP [32]	(93.1 / 93.4)	(78.7 / 94.7)	(73.1 / 78.9)	(70.2 / 94.3)	(84.5 / 94.3)	(68.8 / 91.0)
	InCTRL [93]	(92.5 / 93.4)	(82.2 / 93.1)	(**74.9** / 78.3)	(74.8 / 93.1)	(90.5 / 92.9)	(69.5 / 93.1)
	ResAD [77]	(80.7 / 84.3)	(77.2 / 88.3)	(70.5 / 72.7)	(67.7 / 92.7)	(78.4 / 92.3)	(59.7 / 89.5)
	FoundAD [84]	(89.9 / 90.7)	(85.3 / 96.2)	(71.4 / 80.4)	(70.6 / 95.4)	(91.4 / 91.9)	(62.7 / 87.8)
A1	DRA [18]	(92.2 / 92.9)	(82.3 / 90.3)	(72.3 / 75.3)	(73.4 / 93.9)	(90.5 / 93.5)	(67.2 / 87.5)
	AHL [91]	(91.7 / 93.6)	(82.0 / 91.9)	(70.7 / 76.3)	(73.3 / 94.3)	(90.9 / 93.8)	(**70.4** / 90.4)
N1A1	NAGL [73]	(**95.1** / **96.1**)	(**88.5** / **97.5**)	(71.6 / 75.5)	(**77.1** / **96.5**)	(**92.0** / **95.8**)	(67.7 / **93.7**)
	IDEAL (ours)	(**96.3** / **96.8**)	(**90.8** / **97.8**)	(**75.8** / **82.7**)	(**78.5** / **97.9**)	(**93.4** / **97.0**)	(**74.9** / **94.8**)
N2	RegAD [29]	(92.6 / 93.5)	(87.4 / 92.7)	(73.5 / **82.0**)	(71.9 / 94.5)	(90.8 / 95.6)	(70.7 / 94.1)
	PromptAD [36]	(93.5 / 93.8)	(86.6 / 89.8)	(73.8 / 81.8)	(76.8 / 96.1)	(90.3 / 92.6)	(74.4 / 93.7)
	WinCLIP [32]	(93.4 / 93.6)	(84.3 / 94.8)	(73.5 / 79.3)	(70.8 / 94.7)	(85.0 / 95.5)	(74.8 / 92.7)
	InCTRL [93]	(92.7 / 93.7)	(85.6 / 93.7)	(**75.2** / 78.7)	(75.9 / 93.5)	(93.4 / 93.0)	(**75.0** / 94.1)
	ResAD [77]	(85.7 / 86.3)	(82.4 / 89.8)	(75.0 / 75.1)	(68.4 / 94.4)	(80.1 / 92.9)	(65.4 / 92.2)
	FoundAD [84]	(90.3 / 92.8)	(88.5 / 96.5)	(72.3 / 81.4)	(73.9 / 95.7)	(**94.0** / 93.2)	(72.5 / 91.9)
N2A1	NAGL [73]	(**96.0** / **96.6**)	(**89.1** / **97.5**)	(73.5 / 79.8)	(**80.2** / **97.3**)	(93.5 / **96.3**)	(74.6 / **94.3**)
	IDEAL (ours)	(**97.4** / **97.5**)	(**91.3** / **97.9**)	(**77.6** / **85.3**)	(**81.6** / **98.0**)	(**95.6** / **97.2**)	(**78.5** / **95.6**)
N4	RegAD [29]	(94.0 / 95.1)	(89.5 / 93.3)	(74.6 / **83.7**)	(76.3 / 95.2)	(91.7 / 96.3)	(75.8 / **94.5**)
	PromptAD [36]	(94.9 / 94.6)	(87.1 / 90.6)	(74.1 / 83.4)	(78.3 / 96.5)	(90.8 / 93.2)	(74.5 / 94.0)
	WinCLIP [32]	(93.9 / 93.8)	(84.5 / 95.1)	(73.8 / 79.5)	(71.0 / 95.4)	(87.4 / 95.9)	(75.7 / 93.2)
	InCTRL [93]	(94.3 / 94.0)	(89.5 / 94.2)	(**75.6** / 80.4)	(79.4 / 93.7)	(93.9 / 93.6)	(**77.8** / 93.8)
	ResAD [77]	(87.1 / 86.9)	(83.1 / 90.1)	(75.2 / 76.7)	(70.3 / 95.7)	(81.5 / 95.1)	(69.6 / 93.0)
	FoundAD [84]	(92.9 / 93.2)	(91.0 / 96.7)	(73.2 / 82.8)	(77.6 / 95.8)	(**94.3** / 94.0)	(73.5 / 92.8)
N4A1	NAGL [73]	(**96.9** / **97.0**)	(**91.2** / **97.7**)	(**75.6** / 80.9)	(**81.4** / **97.7**)	(93.7 / **96.5**)	(74.9 / 94.4)
	IDEAL (ours)	(**98.2** / **97.5**)	(**92.7** / **98.0**)	(**79.5** / **86.1**)	(**87.6** / **98.5**)	(**95.9** / **97.5**)	(**80.5** / **95.8**)
N8	RegAD [29]	(95.0 / 95.8)	(**91.9** / 94.7)	(75.8 / 83.9)	(80.8 / 95.7)	(92.2 / **96.5**)	(76.3 / 95.0)
	PromptAD [36]	(96.0 / 94.8)	(87.9 / 92.2)	(74.4 / **84.7**)	(80.6 / 96.5)	(91.2 / 94.1)	(75.6 / 94.6)
	WinCLIP [32]	(95.5 / 94.2)	(86.1 / 95.1)	(74.8 / 80.5)	(74.7 / 95.6)	(89.6 / 96.0)	(76.9 / 93.8)
	InCTRL [93]	(95.8 / 94.4)	(89.8 / 95.3)	(75.9 / 80.6)	(**83.2** / 94.1)	(94.3 / 94.9)	(**79.5** / 94.9)
	ResAD [77]	(89.4 / 87.8)	(83.7 / 90.9)	(75.4 / 79.3)	(75.1 / 96.1)	(83.6 / 95.5)	(71.7 / 94.3)
	FoundAD [84]	(93.2 / 93.5)	(91.3 / 96.8)	(74.3 / 83.0)	(80.3 / 96.2)	(**94.6** / 95.1)	(76.7 / 93.5)
A4	DRA [18]	(93.7 / 94.0)	(86.8 / 90.7)	(75.7 / 77.8)	(81.8 / 94.9)	(93.0 / 93.9)	(73.5 / 89.9)
	AHL [91]	(92.9 / 94.5)	(85.9 / 92.5)	(75.1 / 78.8)	(81.0 / 95.2)	(93.3 / 94.5)	(72.9 / 90.3)
N8A4	NAGL [73]	(**97.2** / **97.1**)	(91.5 / **97.9**)	(**80.5** / 82.3)	(83.1 / **97.9**)	(94.0 / **96.5**)	(75.1 / **95.2**)
	IDEAL (ours)	(**98.6** / **97.8**)	(**93.1** / **98.3**)	(**83.9** / **86.8**)	(**88.0** / **98.7**)	(**96.2** / **97.7**)	(**82.8** / **96.9**)

Competing Methods and Metrics. We compare IDEAL with two groups of closely related methods: (1) *Generalist Methods*, *i.e.*, detectors are training-free or trained once on auxiliary data: based on normal-only references (InCTRL [93], ResAD [77], WinCLIP [32], FoundAD [84]), and normal and anomalous references (NAGL [73]); (2) *Specalist Methods*, *i.e.*, detectors require dataset-specific training: based on normal-only references (RegAD [29], PromptAD [36]), and based on full-shot normal data and abnormal-only references: DRA [18] and AHL [91]. For metrics, following [90, 93], we use Area Under the Receiver Operating Characteristic (AUROC), maximum F1-score at the optimal threshold (F1-Max), and Per-Region Overlap (PRO) to evaluate pixel-level AD; similarly, we use AUROC, F1-Max, and Average Precision (AP) to evaluate image-level AD.

Implementation Details. By default, we adopt a pre-trained ViT-S/14 [20] backbone as the visual encoder $\mathcal{E}(\cdot)$ in our experiments. We freeze the parameters of the visual encoder and only update the parameters of the IDE component. All the training and test images are resized to a 448×448 resolution. AdamW [42] is used as the optimizer, and the initial learning rate is set to 0.001 with a cosine scheduler. The number of training epochs of IDEAL is set to 20 with a batch size of 16 on a single NVIDIA GeForce RTX 3090 GPU. We set the number k of normal neighbors in Eq. (2) as 12,

the number r of retained principal components as 4, α in Eq. (4) as 0.8, M in Eq. (6) as 45, λ_1 and λ_2 in Eq. (7) as 1.0 and 0.8 by default. Following previous studies [91, 93], we set the number of normal references to $L_1 \in [1, 2, 4, 8]$ and the number of abnormal references to $L_2 \in [1, 4]$, with $L_2 = 1$ as the default to simulate scenarios where anomaly data is costly to obtain. For AD methods requiring abnormal references, we sample the abnormal references from a randomly selected anomaly type, and consider two setups: *General Setting*, which utilizes the full test set for evaluation of detecting both seen and unseen anomalies, and *Hard Setting*, which excludes test samples belonging to the selected anomaly type for an exclusive evaluation of detecting unseen anomalies. All results reported are the average of three independent runs (see more details in Appendix C).

5.2 Main Results

Benefits of Abnormal References. Tab. 1 presents comparison results of image-level and pixel-level AUROCs under various few-shot AD settings. Note that all the reported results are dataset-level average results across their respective subsets. Overall, IDEAL achieves superior performance by incorporating both normal and abnormal references. With just one additional abnormal reference `N1A1`, IDEAL yields 96.3% and 96.8% for image and pixel AUROCs on MVTecAD, outperforming state-of-the-art FSAD methods that use normal-only references. The similar improvements are also observed on other AD datasets. These results show the advantage of introducing abnormal examples into the reference set, as they provide discriminative clues about anomaly characteristics that are difficult to obtain from normal-only references. Please see complete results in Appendix H.

General vs. Hard Settings. Tab. 2 reports comparison results of IDEAL and NAGL [73] under the hard setting, where test samples belonging to the anomaly type of abnormal references are excluded from the test set. In contrast to the general setting (Tab. 1), this hard setting evaluates the capability of detecting unseen anomalies. To quantify the hardness/task difficulty of unseen AD, we compute the mask-aware cosine similarity between abnormal references and anomalous test samples, where lower similarity indicates a larger seen-to-unseen anomaly gap (task difficulty values in Tab. 2). In such a hard setting, NAGL [73] exhibits a pronounced performance degradation, as its anomaly scoring relies on the alignment between test query and abnormal references. IDEAL maintains robust performance by leveraging normal-abnormal references to learn intrinsic deviation patterns, enabling generalization to unseen anomalies.

Table 2: **Comparison results (*Hard*) with AUROC** on MVTecAD and MPDD, together with task difficulty values (purple) that measures a seen-to-unseen anomaly similarity.

SETUPS	METHODS	MVTECAD		MPDD	
`N1A1`	NAGL [73]	(93.2 ↓1.9 / 94.0 ↓2.1)	0.824	(63.7 ↓13.4 / 91.8 ↓4.7)	0.543
	IDEAL	(95.8 ↓0.5 / 96.4 ↓0.4)		(74.5 ↓4.0 / 97.2 ↓0.7)	
`N2A1`	NAGL [73]	(93.5 ↓2.5 / 94.8 ↓1.8)	0.833	(71.7 ↓8.5 / 93.0 ↓4.3)	0.515
	IDEAL	(97.0 ↓0.4 / 96.6 ↓0.9)		(77.0 ↓4.6 / 97.4 ↓0.6)	
`N4A1`	NAGL [73]	(94.6 ↓2.3 / 95.0 ↓2.0)	0.839	(75.8 ↓5.6 / 93.8 ↓3.9)	0.501
	IDEAL	(97.3 ↓0.9 / 97.0 ↓0.5)		(84.3 ↓3.3 / 97.5 ↓1.0)	
`N8A4`	NAGL [73]	(95.1 ↓2.1 / 95.4 ↓1.7)	0.835	(76.6 ↓6.5 / 94.2 ↓3.7)	0.513
	IDEAL	(97.6 ↓1.0 / 97.2 ↓0.6)		(84.9 ↓3.1 / 97.8 ↓0.9)	

5.3 Validation Studies

Deviation Analysis. Fig. 3 presents the t-SNE [68] visualization of different deviation representations produced by IDEAL on VisA (trained on MVTecAD). From these results, we can observe: 1) residual deviation representations are highly entangled between normal and abnormal samples, while denoised deviation representations suppress noisy normal variations contained in these residual deviations; 2) after projection onto the learned intrinsic deviation vectors, normal and abnormal samples become clearly separated. These results indicate that IDEAL can effectively extract intrinsic deviations from normality, thereby improving normal-abnormal separability, which in turn enhances detection.

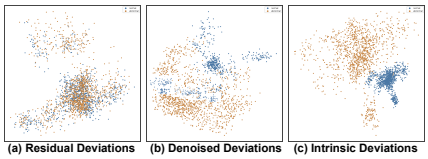

Figure 3: T-SNE visualization for different deviation representations produced by IDEAL on VisA (normal abnormal).

Ablation Studies. We provide an ablation analysis of IDEAL's components in Tab. 3 (**Left**). Comparing first two rows, we can observe that naively matching the normal/abnormal reference to perform anomaly scoring is inherently fragile, as such a strategy fails to cover diverse anomaly types. The third row indicates that our NVE component effectively reduces noisy activations caused by

Table 3: **Left**: Component ablation of IDEAL. The first two rows use nearest neighbor matching to perform anomaly scoring, the third row adds the similarity score between query-to-normal deviations and denoised deviations, the fourth row uses the residual deviations to learn deviation patterns, and the fifth row is our full IDEAL model. **Right**: Efficiency comparisons on VisA (trained on MVTecAD). Note that the time of WinCLIP represents the wall-clock time of the complete pipeline.

N1	A1	NVE	IDE	MVTecAD	VisA	BraTS
	✓			(72.4 / 88.0)	(61.3 / 85.6)	(39.7 / 83.3)
✓	✓			(90.1 / 92.2)	(77.8 / 92.7)	(57.6 / 87.5)
✓	✓	✓		(93.9 / 94.3)	(82.7 / 95.3)	(63.6 / 92.0)
✓	✓		✓	(**94.5** / **96.0**)	(**87.3** / **96.5**)	(**70.1** / **92.9**)
✓	✓	✓	✓	(**96.3** / **96.8**)	(**90.8** / **97.8**)	(**74.9** / **94.8**)

Methods	Param. (M)	Time	FPS	FLOPs$_{\times 10^{10}}$
WinCLIP [32]	117.3	**0.2 H**	1.8	41.52
InCTRL [93]	117.5	0.9 H	1.5	20.47
ResAD [77]	59.31	22.8 H	7.6	15.29
NAGL [73]	**24.45**	0.6 H	**16.8**	**13.98**
IDEAL (ours)	**23.83**	**0.5 H**	**21.9**	**12.70**

nuisance normal variations. The fourth row verifies that IDE is effective even when applied to noisy residual deviations, showing that learning intrinsic deviation vectors itself provides strong anomaly discriminative ability. By further combining NVE and IDE (the fifth row), IDEAL learns intrinsic deviation vectors from denoised deviations, enabling generalization to both seen and unseen anomalies. Appendix D further shows that NVE can be easily incorporated with existing residual-based FSAD methods to enhance detection. In Appendix E and Appendix F, we provide complete ablation studies of IDEAL's key parameters and analyze the effect of loss terms in Eq. (10).

Efficiency Comparison. We evaluate the efficiency from four aspects: model parameters, wall-clock time, inference speed (FPS), computation cost (FLOPs). As shown in Tab. 3 (**Right**), IDEAL has 23.83 million parameters, which is around $5\times$ smaller than InCTRL and WinCLIP. Compared with ResAD, IDEAL requires only 0.5 hours for training, achieving a $45\times$ speedup. In terms of inference speed, IDEAL reaches 21.9 FPS, which is $14\times$ faster than InCTRL and $1.3\times$ faster than NAGL. These results demonstrate the practical efficiency of IDEAL for real-world deployment.

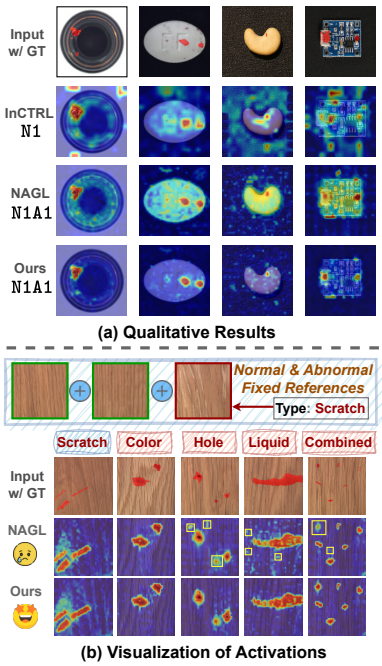

Figure 4: **(a)** Qualitative results of three generalist FSAD methods. **(b)** Visualization of anomalous activations on a new dataset (VisA→MVTecAD).

Qualitative Results. Fig. 4 (a) shows qualitative results of three generalist FSAD methods. Comparing the anomaly score maps, we observe that IDEAL effectively avoids false positives in normal regions and locates anomalies more accurately. Then, Fig. 4 (b) further illustrates the generalizability of IDEAL under limited reference coverage. With a fixed reference set, NAGL is sensitive to the coverage of the given anomalous exemplars and tends to fail when the query exhibits anomaly appearances that are unseen in, or poorly covered by, the references. In contrast, IDEAL learns intrinsic deviation patterns from normal and abnormal references, enabling more robust detection beyond limited references.

Number of Abnormal References. To investigate the demand for abnormal examples, we conduct an ablation study by varying $L_2 \in \{0, 1, 2, 4\}$ while fixing the number L_1 of normal examples. Tab. 4 shows the results of homogeneous (*i.e.*, `homo`, all anomaly references are sampled from one anomaly type) and heterogeneous anomaly references (*i.e.*, `hete`, the references are sampled from as many anomaly types as possible), we can observe that: 1) adding one abnormal reference already provides substantial gains, and 2) increasing L_2 from 1 to 4 yields consistent improvement, but the gains are mostly marginal. The largest gain still comes from adding one abnormal reference.

Effect of Reference Anomaly Types. We conduct an iterative cross-validation of the performance on the anomaly type, *i.e.*, the abnormal reference is drawn from one anomaly type, and the evaluation

Table 4: **Comparison results (*General*) with different** L_2, where `homo` and `hete` represent homogeneous and heterogeneous anomaly references.

SETUPS	DATASETS	N1A0	N1A2	N1A4	N4A0	N4A2	N4A4
homo	MVTecAD	(94.0 / 94.5)	(96.5 / 97.0)	(96.9 / **97.3**)	(94.2 / 94.9)	(**98.2** / **97.6**)	(**98.5** / **97.6**)
	VisA	(84.3 / 94.3)	(91.0 / 97.8)	(91.3 / 97.9)	(88.3 / 94.5)	(**92.9** / **98.2**)	(**93.0** / **98.4**)
	MPDD	(72.6 / 94.8)	(79.4 / **98.1**)	(79.9 / **98.1**)	(83.0 / 96.1)	(**87.8** / **98.5**)	(**87.9** / **98.5**)
hete	MVTecAD	(94.0 / 94.5)	(96.8 / 97.2)	(97.0 / **97.5**)	(94.2 / 94.9)	(**98.4** / **97.5**)	(**98.5** / **97.7**)
	VisA	(84.3 / 94.3)	(91.3 / 97.9)	(91.5 / **98.0**)	(88.3 / 94.5)	(**93.0** / **98.4**)	(**93.4** / **98.4**)
	MPDD	(72.6 / 94.8)	(80.1 / 98.0)	(80.3 / 98.3)	(83.0 / 96.1)	(**87.9** / **98.5**)	(**88.2** / **98.6**)

Table 5: **Cross-validation results (*Hard*) with different reference anomaly types** under the `N1A1` setting, on two MVTecAD datasets (`[wood]` and `[capsule]`).

METHODS	[wood]=color	combined	hole	liquid	scratch
NAGL [73]	(99.8±0.1 / 96.7±0.0)	(99.8±0.0 / 94.6±1.5)	(99.9±0.0 / 95.7±0.1)	(100.0±0.0 / 94.7±0.1)	(99.8±0.1 / 97.3±0.3)
IDEAL (ours)	(99.9±0.0 / 97.0±0.4)	(99.8±0.1 / 97.5±0.0)	(99.9±0.0 / 97.1±0.4)	(100.0±0.0 / 97.5±0.2)	(99.9±0.1 / 97.9±0.1)
METHODS	[capsule]=crack	imprint	poke	squeeze	scratch
NAGL [73]	(82.2±0.8 / 93.1±3.0)	(82.5±4.2 / 95.4±0.6)	(83.7±1.6 / 96.4±0.2)	(85.2±1.0 / 92.6±1.1)	(83.7±0.3 / 94.4±1.1)
IDEAL (ours)	(85.9±0.0 / 97.3±0.5)	(86.1±0.2 / 98.0±0.1)	(85.5±0.4 / 97.5±0.1)	(85.8±0.6 / 97.9±0.2)	(85.4±0.1 / 97.5±0.0)

is done on all remaining anomaly types, while keeping the normal references unchanged. For a dataset that has K (*e.g.*, 5) anomaly types, we perform the cross-validation K times across all anomaly types and report the mean±std results of these K runs on each anomaly type per dataset. Tab. 5 shows that the proposed IDEAL is notably less sensitive to the reference anomaly type than NAGL.

Analysis of Pixel-level Anomaly Masks. The pixel-level masks can provide stronger supervision than normal-only FSAD. IDEAL requires only one abnormal-reference mask per dataset by default ($L_2 = 1$), annotated once and reused for all queries. To isolate this mask benefit, we evaluate an image-level-only variant that replaces ground-truth masks with *pseudo-masks* derived from abnormal-normal feature distances (top-1% set as 1, others set as 0). As shown in Tab. 6, IDEAL remains effective by using pseudo anomaly masks only, having small gaps to the variant using the ground-truth anomaly masks.

Table 6: **Results (*General*) using pseudo** (`Pse`) **/truth anomaly masks**, under the `N1A1` setting.

METHODS	MVTECAD	VISA	MPDD
NAGL [73]	(95.1 / 96.1)	(88.5 / 97.5)	(77.1 / 96.5)
NAGL (Pse)	(94.3 / 95.9)	(86.4 / 95.7)	(75.7 / 96.0)
IDEAL	(96.3 / 96.8)	(90.8 / 97.8)	(78.5 / 97.9)
IDEAL (Pse)	(95.4 / 96.6)	(89.8 / 97.6)	(77.2 / 97.4)

6 Conclusion

In this paper, we consider a practical few-shot anomaly detection (FSAD) setting, discriminative FSAD, where a few normal and anomalous examples are available as references at inference. To effectively leverage discriminative clues of abnormal references relative to normal references, we propose an Intrinsic Deviation Learning (IDEAL) framework to learn intrinsic deviation patterns characterizing generalizable abnormality as deviations from normality, for detecting both seen and unseen anomalies. At inference, anomalies are detected by measuring query-to-normal deviations preserved after projection onto learned intrinsic deviation vectors. Extensive experiments validate the effectiveness and efficiency of our proposed IDEAL.

Acknowledgments

This work is supported by A*STAR under its MTC YIRG Grant (M24N8c0103), the National Research Foundation, Singapore and CyberSG R&D Programme Office under its Translation and Innovation Grants (CRPO-GC4-SMU-001), the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 Grant (24-SIS-SMU-008), and the Lee Kong Chian Fellowship (T050273).

References

[1] Hervé Abdi and Lynne J Williams. Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4):433–459, 2010. 5

[2] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *ACCV*, pages 622–637, 2018. 3

[3] Abdulatif Alabdulatif, Heshan Kumarage, Ibrahim Khalil, and Xun Yi. Privacy-preserving anomaly detection in cloud with lightweight homomorphic encryption. *Journal of Computer and System Sciences*, 90:28–45, 2017. 3

[4] Jinan Bao, Hanshi Sun, Hanqiu Deng, Yinsheng He, Zhaoxiang Zhang, and Xingyu Li. Bmad: Benchmarks for medical anomaly detection. In *CVPR*, pages 4042–4053, 2024. 6, 20, 21

[5] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine Learning*, 79(1):151–175, 2010. 15, 16

[6] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. In *NeurIPS*, pages 137–144, 2006. 15, 16

[7] Liron Bergman and Yedid Hoshen. Classification-based anomaly detection for general data. *arXiv preprint arXiv:2005.02359*, 2020. 3

[8] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, pages 9592–9600, 2019. 6, 20, 21

[9] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *CVPR*, pages 4183–4192, 2020. 3

[10] Tri Cao, Jiawen Zhu, and Guansong Pang. Anomaly detection under distribution shift. In *ICCV*, pages 6511–6523, 2023. 3

[11] Yunkang Cao, Xiaohao Xu, Jiangning Zhang, Yuqi Cheng, Xiaonan Huang, Guansong Pang, and Weiming Shen. A survey on visual anomaly detection: Challenge, approach, and prospect. *arXiv preprint arXiv:2401.16402*, 2024. 1, 2

[12] Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019. 2

[13] Yuanhong Chen, Yu Tian, Guansong Pang, and Gustavo Carneiro. Deep one-class classification via interpolated gaussian descriptor. In *AAAI*, pages 383–392, 2022. 3

[14] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020. 1, 3

[15] Simon Damm, Mike Laszkiewicz, Johannes Lederer, and Asja Fischer. Anomalydino: Boosting patch-based few-shot anomaly detection with dinov2. In *WACV*, pages 1319–1329, 2025. 3

[16] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *ICPR*, pages 475–489, 2021. 1, 3

[17] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, pages 9737–9746, 2022. 3

[18] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *CVPR*, pages 7388–7398, 2022. 1, 3, 7, 21, 22, 25, 26, 27, 28, 29, 30, 31, 32

[19] Le Dong, Qinzhong Tan, Chunlei Li, Jingliang Hu, Yilei Shi, Weisheng Dong, Xiao Xiang Zhu, and Lichao Mou. Dual distillation for few-shot anomaly detection. In *ICLR*, 2026. 3

[20] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, 2021. 7

[21] Fang Fang, Kaushik Dutta, and Anindya Datta. Domain adaptation for sentiment classification in light of multiple sources. *INFORMS Journal on Computing*, 26(3):586–598, 2014. 15

[22] Zheng Fang, Xiaoyang Wang, Haocheng Li, Jiejie Liu, Qiugui Hu, and Jimin Xiao. Fastrecon: Few-shot industrial anomaly detection via fast feature reconstruction. In *ICCV*, pages 17481–17490, 2023. 1, 3

[23] Tharindu Fernando, Harshala Gammulle, Simon Denman, Sridha Sridharan, and Clinton Fookes. Deep learning for medical anomaly detection–a survey. *ACM Computing Surveys*, 54(7):1–37, 2021. 1

[24] Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Hao Li, Ming Tang, and Jinqiao Wang. Filo: Zero-shot anomaly detection by fine-grained description and high-quality localization. In *ACM MM*, pages 2041–2049, 2024. 3

[25] Jia Guo, Shuai Lu, Weihang Zhang, Fang Chen, Huiqi Li, and Hongen Liao. Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In *CVPR*, pages 20405–20415, 2025. 3

[26] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *AAAI*, pages 8472–8480, 2024. 1

[27] Jinlei Hou, Yingying Zhang, Qiaoyong Zhong, Di Xie, Shiliang Pu, and Hong Zhou. Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection. In *ICCV*, pages 8791–8800, 2021. 3

[28] Junjie Hu, Yuanyuan Chen, and Zhang Yi. Automated segmentation of macular edema in oct using deep neural networks. *Medical Image Analysis*, 55:216–227, 2019. 6, 20, 21

[29] Chaoqin Huang, Haoyan Guan, Aofan Jiang, Ya Zhang, Michael Spratling, and Yan-Feng Wang. Registration based few-shot anomaly detection. In *ECCV*, pages 303–319, 2022. 1, 3, 7, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32

[30] Chaoqin Huang, Aofan Jiang, Jinghao Feng, Ya Zhang, Xinchao Wang, and Yanfeng Wang. Adapting visual-language models for generalizable anomaly detection in medical images. In *CVPR*, pages 11375–11385, 2024. 1

[31] Jaehui Hwang, Junghyuk Lee, and Jong-Seok Lee. Anomaly score: Evaluating generative models and individual generated images based on complexity and vulnerability. In *CVPR*, pages 8754–8763, 2024. 6

[32] Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. Winclip: Zero-/few-shot anomaly classification and segmentation. In *CVPR*, pages 19606–19616, 2023. 1, 3, 7, 9, 21, 25, 26, 27, 28, 29, 30, 31, 32

[33] Stepan Jezek, Martin Jonak, Radim Burget, Pavel Dvorak, and Milos Skotak. Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In *International Congress on Ultra Modern Telecommunications and Control Systems and Workshops*, pages 66–71, 2021. 6, 20, 21

[34] Aodong Li, Chen Qiu, Marius Kloft, Padhraic Smyth, Maja Rudolph, and Stephan Mandt. Zero-shot anomaly detection via batch normalization. In *NeurIPS*, pages 40963–40993, 2023. 3

[35] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *CVPR*, pages 9664–9674, 2021. 3

[36] Xiaofan Li, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yuan Xie, and Lizhuang Ma. Promptad: Learning prompts with only normal samples for few-shot anomaly detection. In *CVPR*, pages 16838–16848, 2024. 1, 3, 7, 21, 22, 25, 26, 27, 28, 29, 30, 31, 32

[37] Yanhui Li, Yunkang Cao, Chengliang Liu, Yuan Xiong, Xinghui Dong, and Chao Huang. Iad-r1: Reinforcing consistent reasoning in industrial anomaly detection. In *AAAI*, pages 6583–6591, 2026. 1

[38] Jingyi Liao, Xun Xu, Manh Cuong Nguyen, Adam Goodge, and Chuan Sheng Foo. Coft-ad: Contrastive fine-tuning for few-shot anomaly detection. *IEEE Transactions on Image Processing*, 33:2090–2103, 2024. 3

[39] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, pages 2980–2988, 2017. 6

[40] Wenrui Liu, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Diversity-measurable anomaly detection. In *CVPR*, pages 12147–12156, 2023. 3

[41] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *CVPR*, pages 20402–20411, 2023. 3

[42] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *ICLR*, 2019. 7

[43] Wei Luo, Yunkang Cao, Haiming Yao, Xiaotian Zhang, Jianan Lou, Yuqi Cheng, Weiming Shen, and Wenyong Yu. Exploring intrinsic normal prototypes within a single image for universal anomaly detection. In *CVPR*, pages 9974–9983, 2025. 3

[44] Wenxi Lv, Qinliang Su, and Wenchao Xu. One-for-all few-shot anomaly detection via instance-induced prompt learning. In *ICLR*, 2025. 1, 3

[45] Zhisheng Lv, Jianfeng Zhang, Songlei Jian, Chenlin Huang, Hongguang Zhang, Guansong Pang, and Zhong Liu. Metacan: Improving generalizability of few-shot anomaly detection with meta-learning. In *CIKM*, pages 2032–2041, 2025. 1, 3

[46] Wenxin Ma, Xu Zhang, Qingsong Yao, Fenghe Tang, Chenxu Wu, Yingtai Li, Rui Yan, Zihang Jiang, and S. Kevin Zhou. Aa-clip: Enhancing zero-shot anomaly detection via anomaly-aware clip. In *CVPR*, pages 4744–4754, 2025. 3

[47] Shie Mannor, Dori Peleg, and Reuven Rubinstein. The cross entropy method for classification. In *ICML*, pages 561–568, 2005. 6

[48] Yishay Mansour, Mehryar Mohri, and Afshin Rostamizadeh. Domain adaptation with multiple sources. In *NeurIPS*, pages 1041–1048, 2008. 15, 18, 19

[49] Rudolf Mayer, Markus Hittmeir, and Andreas Ekelhart. Privacy-preserving anomaly detection using synthetic data. In *IFIP Annual Conference on Data and Applications Security and Privacy*, pages 195–207. Springer, 2020. 3

[50] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024, 2015. 6, 20, 21

[51] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *International Conference on 3D Vision*, pages 565–571, 2016. 6

[52] Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, and Gian Luca Foresti. Vt-adl: A vision transformer network for image anomaly detection and localization. In *International Symposium on Industrial Electronics*, pages 01–06, 2021. 6, 20, 21

[53] Guansong Pang, Longbing Cao, Ling Chen, and Huan Liu. Learning representations of ultrahigh-dimensional data for random distance-based outlier detection. In *ACM SIGKDD*, pages 2041–2050, 2018. 3

[54] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM Computing Surveys*, 54(2):1–38, 2021. 1, 2

[55] Guansong Pang, Chunhua Shen, and Anton Van Den Hengel. Deep anomaly detection with deviation networks. In *ACM SIGKDD*, pages 353–362, 2019. 1, 3

[56] Hyunjong Park, Jongyoun Noh, and Bumsub Ham. Learning memory-guided normality for anomaly detection. In *CVPR*, pages 14372–14381, 2020. 3

[57] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, pages 14318–14328, 2022. 1, 3

[58] Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. Deep semi-supervised anomaly detection. In *ICLR*, 2020. 3

[59] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, and Hamid R Rabiee. Multiresolution knowledge distillation for anomaly detection. In *CVPR*, pages 14902–14912, 2021. 3

[60] Durgesh Samariya and Amit Thakkar. A comprehensive survey of anomaly detection algorithms. *Annals of Data Science*, 10(3):829–850, 2023. 2

[61] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54:30–44, 2019. 3

[62] Javier Silvestre-Blanes, Teresa Albero-Albero, Ignacio Miralles, Rubén Pérez-Llorens, and Jorge Moreno. A public fabric database for defect detection methods and results. *Autex Research Journal*, 19(4):363–374, 2019. 6, 20, 21

[63] Qian Sun, Rita Chattopadhyay, Sethuraman Panchanathan, and Jieping Ye. A two-stage weighting framework for multi-source domain adaptation. In *NeurIPS*, pages 505–513, 2011. 16

[64] Fenfang Tao, Guo-Sen Xie, Fang Zhao, and Xiangbo Shu. Kernel-aware graph prompt learning for few-shot anomaly detection. In *AAAI*, pages 7347–7355, 2025. 3

[65] Tareq Tayeh, Sulaiman Aburakhia, Ryan Myers, and Abdallah Shami. Distance-based anomaly detection for industrial surfaces using triplet networks. In *IEMCON*, pages 0372–0377. IEEE, 2020. 3

[66] Long Tian, Hongyi Zhao, Ruiying Lu, Rongrong Wang, Yujie Wu, Liming Wang, Xiongpeng He, and Xiyang Liu. Foct: Few-shot industrial anomaly detection with foreground-aware online conditional transport. In *ACM MM*, pages 6241–6249, 2024. 3

[67] Michael E Tipping and Christopher M Bishop. Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(3):611–622, 1999. 5

[68] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(11):2579–2605, 2008. 8

[69] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *NeurIPS*, pages 5998–6008, 2017. 5

[70] Fuyun Wang, Tong Zhang, Yuanzhi Wang, Yide Qiu, Xin Liu, Xu Guo, and Zhen Cui. Distribution prototype diffusion learning for open-set supervised anomaly detection. In *CVPR*, pages 20416–20426, 2025. 3

[71] Jue Wang and Anoop Cherian. Gods: Generalized one-class discriminative subspaces for anomaly detection. In *ICCV*, pages 8201–8211, 2019. 1

[72] Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *CVPR*, pages 8032–8041, 2023. 1

[73] Yuexin Wang, Xiaolei Wang, Yizheng Gong, and Jimin Xiao. Normal-abnormal guided generalist anomaly detection. In *NeurIPS*, 2025. 3, 6, 7, 8, 9, 10, 21, 23, 25, 26, 27, 28, 29, 30, 31, 32

[74] Ze Wang, Yipin Zhou, Rui Wang, Tsung-Yu Lin, Ashish Shah, and Ser Nam Lim. Few-shot fast-adaptive anomaly detection. In *NeurIPS*, pages 4957–4970, 2022. 1, 3

[75] Guoyang Xie, Jinbao Wang, Jiaqi Liu, Yaochu Jin, and Feng Zheng. Pushing the limits of fewshot anomaly detection in industry vision: Graphcore. In *ICLR*, 2023. 1, 3

[76] Jiacong Xu, Shao-Yuan Lo, Bardia Safaei, Vishal M Patel, and Isht Dwivedi. Towards zero-shot anomaly detection and reasoning with multimodal large language models. In *CVPR*, pages 20370–20382, 2025. 3

[77] Xincheng Yao, Zixin Chen, Chao Gao, Guangtao Zhai, and Chongyang Zhang. Resad: A simple framework for class generalizable anomaly detection. In *NeurIPS*, pages 125287–125311, 2024. 1, 3, 6, 7, 9, 21, 23, 25, 26, 27, 28, 29, 30, 31, 32

[78] Xincheng Yao, Ruoqi Li, Zefeng Qian, Lu Wang, and Chongyang Zhang. Hierarchical gaussian mixture normalizing flow modeling for unified anomaly detection. In *ECCV*, pages 92–108, 2024. 1

[79] Xincheng Yao, Ruoqi Li, Jing Zhang, Jun Sun, and Chongyang Zhang. Explicit boundary guided semi-push-pull contrastive learning for supervised anomaly detection. In *CVPR*, pages 24490–24499, 2023. 1

[80] Xincheng Yao, Chongyang Zhang, Ruoqi Li, Jun Sun, and Zhenyu Liu. One-for-all: Proposal masked cross-class anomaly detection. In *AAAI*, pages 4792–4800, 2023. 1

[81] Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *ACCV*, 2020. 3

[82] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. In *NeurIPS*, pages 4571–4584, 2022. 1

[83] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, pages 8330–8339, 2021. 3

[84] Guangyao Zhai, Yue Zhou, Xinyan Deng, Lars Heckler-Kram, Nassir Navab, and Benjamin Busam. Foundation visual encoders are secretly few-shot anomaly detectors. In *ICLR*, 2026. 1, 3, 7, 21, 25, 26, 27, 28, 29, 30, 31, 32

[85] Jianpeng Zhang, Yutong Xie, Guansong Pang, Zhibin Liao, Johan Verjans, Wenxing Li, Zongji Sun, Jian He, Yi Li, Chunhua Shen, et al. Viral pneumonia screening on chest x-rays using confidence-aware anomaly detection. *IEEE transactions on medical imaging*, 40(3):879–890, 2020. 1

[86] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *CVPR*, pages 16699–16708, 2024. 3

[87] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *CVPR*, pages 3914–3923, 2023. 3

[88] Yabin Zhang, Bin Deng, Hui Tang, Lei Zhang, and Kui Jia. Unsupervised multi-class domain adaptation: Theory, algorithms, and practice. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2775–2792, 2020. 16

[89] Yuchen Zhang, Tianle Liu, Mingsheng Long, and Michael Jordan. Bridging theory and algorithm for domain adaptation. In *ICML*, pages 7404–7413, 2019. 16

[90] Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. In *ICLR*, 2024. 3, 7

[91] Jiawen Zhu, Choubo Ding, Yu Tian, and Guansong Pang. Anomaly heterogeneity learning for open-set supervised anomaly detection. In *CVPR*, pages 17616–17626, 2024. 3, 7, 8, 21, 22, 25, 26, 27, 28, 29, 30, 31, 32

[92] Jiawen Zhu, Yew-Soon Ong, Chunhua Shen, and Guansong Pang. Fine-grained abnormality prompt learning for zero-shot anomaly detection. In *ICCV*, pages 22241–22251, 2025. 3

[93] Jiawen Zhu and Guansong Pang. Toward generalist anomaly detection via in-context residual learning with few-shot sample prompts. In *CVPR*, pages 17826–17836, 2024. 1, 3, 6, 7, 8, 9, 21, 22, 23, 25, 26, 27, 28, 29, 30, 31, 32

[94] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *ECCV*, pages 392–408, 2022. 6, 20, 21

A Algorithm Pseudo-code Flow

In this section, we describe the pseudo-code of IDEAL in Algorithm 1: 1) *Training on Auxiliary Datasets*: IDEAL is trained on auxiliary datasets in an episodic manner, where each training episode is constructed to simulate the few-shot inference scenario, consisting of a query image together with limited normal and abnormal references sampled from the source dataset. 2) *Inference*: IDEAL is evaluated on the target dataset without further training. For each target dataset, a few fixed normal and abnormal references are provided together with a test query. Based on learned intrinsic deviation vectors, IDEAL produces deviation-guided anomaly scores for each test query.

Algorithm 1 **IDEAL**: Intrinsic Deviation Learning for Discriminative FSAD

/* Training on Auxiliary Datasets */
$\mathcal{D}_{tr}^{q} \Leftarrow \mathcal{D}_{tr}$ // randomly select training queries from auxiliary datasets
for each training query $\mathbf{x}^q$ from $\mathcal{D}_{tr}^{q}$ **do**
  $\mathcal{S}^n, \mathcal{S}^a \Leftarrow \mathcal{D}_{tr}$ // randomly sample normal and abnormal references
  $\mathbf{E} = \{\mathbf{x}^q, \mathcal{S}^n, \mathcal{S}^a\}$ // construct training episode
  $\mathcal{F}^q, \mathcal{F}^n, \mathcal{F}^a \leftarrow \mathcal{E}(\mathbf{x}^q), \mathcal{E}(\mathcal{S}^n), \mathcal{E}(\mathcal{S}^a)$ // extracting features
  /* Normal Variation Eraser (Sec. 4.1) */
  $\mathcal{F}^a_{\texttt{res}} = \{f^a_{i,\texttt{res}}\}_{i=1}^{L_2 \cdot N} \leftarrow$ Eq. (1) // residual deviations
  $\mathcal{F}^a_{\texttt{den}} = \{f^a_{i,\texttt{den}}\}_{i=1}^{L_2 \cdot N} \leftarrow$ Eq. (4) // denoised deviations
  /* Intrinsic Deviation Encoder (Sec. 4.2) */
  $\mathcal{T} \in \mathbb{R}^{M \times \mathbb{C}}$ // learnable vectors
  $\mathcal{T}^* = \{t_m\}_{1 \le m \le M} \leftarrow$ Eq. (6) // intrinsic deviation vectors
  /* Anomaly Scoring with Projection (Sec. 4.3) */
  $\mathcal{F}^q_{\texttt{den}} = \{f^q_{i,\texttt{den}}\}_{i=1}^{N} \leftarrow$ Eq. (4) // query's denoised deviations
  $\tilde{\mathbf{f}}^q_{i,\texttt{den}} = \texttt{Proj}(f^q_{i,\texttt{den}}) \leftarrow$ Eq. (8) // projection
  $A^q, \mathcal{A}^q \leftarrow$ Eq. (9) // anomaly scoring
  $\mathcal{L}_{Dual}(\mathcal{F}^a_{\texttt{den}}, \mathcal{T}^*) \leftarrow$ Eq. (7) // dual-branch loss
  $\mathcal{L}_{Focal}(A^q, \mathcal{M}^q) \leftarrow$ Eq. (10) // Focal loss
  $\mathcal{L}_{Dice}(A^q, \mathcal{M}^q) \leftarrow$ Eq. (10) // Dice loss
  $\mathcal{L}_{BCE}(\mathcal{A}^q, \mathbf{y}^q) \leftarrow$ Eq. (10) // BCE loss
  $\mathcal{L}_{train} = \mathcal{L}_{Focal} + \mathcal{L}_{Dice} + \mathcal{L}_{BCE} + \mathcal{L}_{Dual} \leftarrow$ Eq. (10) // training objective
  $\Phi \leftarrow \Phi - \eta \nabla \mathcal{L}_{train}(\Phi; \mathbf{E})$ // update model
end
/* Inference on Target Dataset */
$\mathcal{S}^n_c, \mathcal{S}^a_c$ // fixed normal and abnormal references for each target dataset c
for each test query $\mathbf{x}^q_c$ from the target dataset c **do**
  $\mathbf{E} = \{\mathbf{x}^q_c, \mathcal{S}^n_c, \mathcal{S}^a_c\}$ // construct inference episode
  $\mathcal{F}^q_c, \mathcal{F}^n_c, \mathcal{F}^a_c \leftarrow \mathcal{E}(\mathbf{x}^q_c), \mathcal{E}(\mathcal{S}^n_c), \mathcal{E}(\mathcal{S}^a_c)$ // extracting features
  $\mathcal{T}^* = \{t_m\}_{1 \le m \le M}$ // learned intrinsic deviation vectors
  $\mathcal{F}^q_{c,\texttt{den}} \leftarrow$ Eq. (4) // query's denoised deviations
  $\tilde{\mathbf{f}}^q_{c,i,\texttt{den}} \leftarrow$ Eq. (8) // projection
  $A^q_c \leftarrow$ Eq. (9) // pixel-level anomaly score
  $\mathcal{A}^q_c \leftarrow$ Eq. (9) // image-level anomaly score
end

B Generalization Analysis of IDEAL

In this section, as in standard theoretical treatments of domain adaptation, we analyze the generalization of IDEAL in two steps because target data are unavailable during training. We begin with a generalization bound on source datasets (Sec. B.3), and subsequently extend this bound to the target dataset (Sec. B.4) based on the multi-source domain adaptation theory developed in [5, 6, 21, 48].

B.1 Preliminaries

Here, we introduce some additional variables to better represent the process of IDEAL. We begin by formalizing the source training process at the episode level. Suppose there are K labeled source domains $\{S_k\}_{k=1}^K$, and $\{\mathbb{P}_{S_1}, \ldots, \mathbb{P}_{S_K}\}$ denote corresponding source episode distributions. Each input episode is of the form $\mathbf{E} = \{\mathbf{x}^q, \mathcal{S}^n, \mathcal{S}^a\}$, where the query $\mathbf{x}^q$ with a ground-truth mask $\mathcal{M}^q$ and a label $\mathbf{y}^q$, $\mathcal{S}^n$ and $\mathcal{S}^a$ are the normal and abnormal reference set. Let $\mathcal{L}(\Phi; \mathbf{E})$ represent the episode-wise loss induced by IDEAL, *i.e.*, the training objective in Eq. (10), and rescale it to $[0, 1]$. Denote the model Φ as the hypothesis from $\mathcal{H}$ and d be the VC-dimension of $\mathcal{H}$. For any source weighting vector $\beta = (\beta_1, \ldots, \beta_K) \in \Delta_K$, following [5, 63], define the weighted source risk is:

$$\epsilon_\beta(\Phi) = \sum_{k=1}^{K} \beta_k \mathbb{E}_{\mathbf{E} \sim \mathbb{P}_{S_k}} [\mathcal{L}(\Phi; \mathbf{E})]. \tag{A1}$$

Suppose that, for the k-th source domain, we observe n_k training episodes as $\{\mathbf{E}_{k,i}\}_{i=1}^{n_k}$, and let the total number as $n = \sum_{k=1}^K n_k$. Thus, the corresponding empirical weighted source risk is:

$$\hat{\epsilon}_\beta(\Phi) = \sum_{k=1}^{K} \beta_k \frac{1}{n_k} \sum_{i=1}^{n_k} \mathcal{L}(\Phi; \mathbf{E}_{k,i}). \tag{A2}$$

Based on learned intrinsic deviation patterns of IDEAL, the episode-loss class can be decomposed as $\mathcal{G} = \cup_{m=1}^M \mathcal{G}_m$, where M is the number of intrinsic deviation patterns and $\mathcal{G}_m$ is the episode-loss subclass associated with the m-th learned deviation pattern. Equivalently, $\mathcal{G}$ is the episode-loss class induced by IDEAL over all deviation patterns: $\mathcal{G} = \{\mathbf{E} \mapsto \mathcal{L}(\Phi; \mathbf{E}) : \Phi \in \mathcal{H}\}$.

B.2 Assumptions

Let $\mathcal{L}(\Phi; \mathbf{E})$ represent the episode-wise loss induced by IDEAL, and let $\mathcal{G}$ be the induced loss class, our generalization analysis follows standard settings in general domain adaptation works [6, 88, 89] and is based on the following assumptions:

Assumption B.1 (Episode Sampling within Source Domains)**.** *For any labeled source domain S_k, the training episodes $\{\mathbf{E}_{k,i}\}_{i=1}^{n_k}$ are drawn i.i.d. from the underlying distribution $\mathbb{P}_{S_k}$. Moreover, the episode sets from different source domains $\{S_1, \ldots, S_K\}$ are mutually independent.*

Assumption B.2 (Bounded Episode Loss)**.** *There exists a constant $B > 0$ such that for all $\Phi \in \mathcal{H}$ and all episodes $\mathbf{E}$, have $0 \leq \mathcal{L}(\Phi; \mathbf{E}) \leq B$. Without loss of generality, we normalize $B = 1$.*

Assumption B.3 (Hypothesis Capacity)**.** *The hypothesis class $\mathcal{H}$ has finite VC-dimension d. Consequently, the induced episode-loss class $\mathcal{G}$ has finite pseudo-dimension.*

Assumption B.4 (Lipschitz Continuity)**.** *Each real-valued loss function $\mathcal{L}(\Phi; \mathbf{E})$ is L-Lipschitz continuous with respect to the model output.*

B.3 Proof of Source-Side Generalization Bound

Theorem B.1 (Source-Side Generalization Bound of IDEAL)**.** *Consider there are K labeled source domains. Let $\mathbb{P}_{S_1}, \ldots, \mathbb{P}_{S_K}$ be the corresponding source episode distributions. Denote the model Φ as the hypothesis from $\mathcal{H}$ and d be the VC-dimension of $\mathcal{H}$. The total number of episodes over all K source domains is n. Then, with probability at least $1 - \delta$,*

$$\sup_{\Phi \in \mathcal{H}} |\epsilon_\beta(\Phi) - \hat{\epsilon}_\beta(\Phi)| \leq \sqrt{\frac{1}{2}\left(\sum_{k=1}^{K} \frac{\beta_k^2}{n_k}\right) \log \frac{M}{\delta} + 2 \sum_{k=1}^{K} \beta_k \sqrt{\frac{d}{n_k} \log \frac{e n_k}{d}}}. \tag{A3}$$

Theorem B.1 indicates that the source-side generalization bound of IDEAL is associated with the number M of intrinsic deviation patterns and the weighting vector β across K source domains. The bound becomes tighter by using appropriate $\Phi \in \mathcal{H}$ and the number M of intrinsic deviation patterns.

Proof of Theorem B.1. For each learned intrinsic deviation pattern of IDEAL indexed by m, based on the episode-loss subclass $\mathcal{G}_m$, define the corresponding deviation quantity as:

$$\Gamma_m = \sup_{g \in \mathcal{G}_m} \left| \sum_{k=1}^{K} \beta_k \mathbb{E}_{\mathbf{E} \sim \mathbb{P}_{S_k}} [g(\mathbf{E})] - \sum_{k=1}^{K} \beta_k \frac{1}{n_k} \sum_{i=1}^{n_k} g(\mathbf{E}_{k,i}) \right|, \tag{A4}$$

since the loss class $\mathcal{G} = \cup_{m=1}^{M} \mathcal{G}_m$ is the union of the subclasses, we can further get:

$$\sup_{\Phi \in \mathcal{H}} |\epsilon_\beta(\Phi) - \hat{\epsilon}_\beta(\Phi)| \leq \max_{1 \leq m \leq M} \Gamma_m. \tag{A5}$$

We first bound the deviation term under Assumption B.1 and Assumption B.2. Replacing a single episode $\mathbf{E}_{k,i}$ changes the weighted empirical operator by at most β_k / n_k. Therefore, by McDiarmid's inequality, for any fixed m and any variable $\xi > 0$, we have:

$$\Pr(\Gamma_m - \mathbb{E}[\Gamma_m] > \xi) \leq \exp\left(-\frac{2\xi^2}{\sum_{k=1}^{K} \beta_k^2 / n_k}\right), \tag{A6}$$

which means that with probability at least $1 - \exp(-\frac{2\xi^2}{\sum_{k=1}^{K} \beta_k^2 / n_k})$,

$$\Gamma_m - \mathbb{E}[\Gamma_m] \leq \xi. \tag{A7}$$

Next, applying the union bound over the M intrinsic deviation patterns, we can get:

$$\Pr\left(\max_{1 \leq m \leq M} \Gamma_m > \max_{1 \leq m \leq M} \mathbb{E}[\Gamma_m] + \xi\right) \leq M \exp\left(-\frac{2\xi^2}{\sum_{k=1}^{K} \beta_k^2 / n_k}\right). \tag{A8}$$

Setting the right-hand side of Eq. (A8) to δ, *i.e.*, $\delta = M \exp(-\frac{2\xi^2}{\sum_{k=1}^{K} \beta_k^2 / n_k})$. Then, we can obtain that with probability at least $1 - \delta$,

$$\max_{1 \leq m \leq M} \Gamma_m \leq \max_{1 \leq m \leq M} \mathbb{E}[\Gamma_m] + \sqrt{\frac{1}{2}\left(\sum_{k=1}^{K} \frac{\beta_k^2}{n_k}\right) \log \frac{M}{\delta}}. \tag{A9}$$

Considering the expectation term of Eq. (A9), by using Rademacher complexity, for each subclass $\mathcal{G}_m$, the following holds:

$$\mathbb{E}[\Gamma_m] \leq 2 \sum_{k=1}^{K} \beta_k \mathbb{E}_\sigma \left[\sup_{g \in \mathcal{G}_m} \frac{1}{n_k} \sum_{i=1}^{n_k} \sigma_i g(\mathbf{E}_{k,i})\right], \tag{A10}$$

where $\sigma = (\sigma_1, \ldots, \sigma_{n_k})$ represent the Rademacher random variables. Eq. (A10) denotes empirical Rademacher complexity of $\mathcal{G}_m$ on the source domain S_k. By Assumption B.3, the induced loss class has finite capacity controlled by d. Hence, using the standard complexity bound, we have:

$$\mathbb{E}_\sigma \left[\sup_{g \in \mathcal{G}_m} \frac{1}{n_k} \sum_{i=1}^{n_k} \sigma_i g(\mathbf{E}_{k,i})\right] \leq \sqrt{\frac{d}{n_k} \log \frac{e n_k}{d}}. \tag{A11}$$

Based on Eq. (A10) and Eq. (A11), we can get:

$$\mathbb{E}[\Gamma_m] \leq 2 \sum_{k=1}^{K} \beta_k \sqrt{\frac{d}{n_k} \log \frac{e n_k}{d}}, \tag{A12}$$

since this upper bound is independent of m, we can further get:

$$\max_{1 \leq m \leq M} \mathbb{E}[\Gamma_m] \leq 2 \sum_{k=1}^{K} \beta_k \sqrt{\frac{d}{n_k} \log \frac{e n_k}{d}}. \tag{A13}$$

Combining the concentration bound Eq. (A9) and the capacity bound Eq. (A13), with probability at least $1 - \delta$, we have:

$$\sup_{\Phi \in \mathcal{H}} |\epsilon_\beta(\Phi) - \hat{\epsilon}_\beta(\Phi)| \leq \sqrt{\frac{1}{2}\left(\sum_{k=1}^{K} \frac{\beta_k^2}{n_k}\right) \log \frac{M}{\delta}} + 2 \sum_{k=1}^{K} \beta_k \sqrt{\frac{d}{n_k} \log \frac{e n_k}{d}}. \tag{A14}$$

Eq. (A14) indicates that the source-side generalization bound of IDEAL is associated with the number M of intrinsic deviation patterns and the weighting vector β across K source domains. □

B.4 Proof of Source-Target Generalization Bound

Since the target data are unavailable during training, we further extend this source-side generalization bound of IDEAL in Eq. (A14) to the target domain via multi-source domain adaptation [48].

Let $\mathbb{P}_T$ denote the target episode distribution, and each target episode $\mathbf{E} = \{\mathbf{x}_T^q, \mathcal{S}_T^n, \mathcal{S}_T^a\}$ has the same form as the source episode, where $\mathbf{x}_T^q$ is the target sample, and $\mathcal{S}_T^n$ and $\mathcal{S}_T^a$ are the corresponding normal and abnormal reference sets from the target domain T, respectively. Further, define the target risk of IDEAL by:

$$\epsilon_T(\Phi) = \mathbb{E}_{\mathbf{E}\sim\mathbb{P}_T}[\mathcal{L}(\Phi; \mathbf{E})]. \tag{A15}$$

Based on the source weighting vector $\beta = (\beta_1, \ldots, \beta_K) \in \Delta_K$, define the weighted source mixture distribution as:

$$\mathbb{P}_\beta = \sum_{k=1}^{K} \beta_k \mathbb{P}_{S_k}, \tag{A16}$$

where $\mathbb{P}_{S_k}$ means the source episode distribution on the source domain S_k.

Following the assumption that the target distribution is some mixture of the source distributions [48], let ℓ be a bounded symmetric loss that measures the disagreement between two hypotheses on an episode. For any episode distribution $\mathbb{P}$, define the pairwise expected disagreement as:

$$\epsilon_{\mathbb{P}}^{\ell}(\Phi, \Phi') = \mathbb{E}_{\mathbf{E}\sim\mathbb{P}}[\ell(\Phi(\mathbf{E}), \Phi'(\mathbf{E}))]. \tag{A17}$$

Then, the discrepancy between the target distribution and the weighted source mixture is:

$$\mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta) = \sup_{\Phi,\Phi'\in\mathcal{H}} \left| \epsilon_{\mathbb{P}_T}^{\ell}(\Phi, \Phi') - \epsilon_{\mathbb{P}_\beta}^{\ell}(\Phi, \Phi') \right|. \tag{A18}$$

Further, define the weighted joint optimal risk as:

$$\Omega_\beta = \inf_{\Phi\in\mathcal{H}} (\epsilon_T(\Phi) + \epsilon_\beta(\Phi)), \tag{A19}$$

where $\epsilon_\beta(\Phi)$ in Eq. (A1) means the weighted source risk.

For convenience, let the source-side generalization bound in Theorem B.1 be denoted by:

$$\mathfrak{B}_{src}(\beta, n, M, d, \delta) = \sqrt{\frac{1}{2}\left(\sum_{k=1}^{K} \frac{\beta_k^2}{n_k}\right) \log \frac{M}{\delta}} + 2\sum_{k=1}^{K} \beta_k \sqrt{\frac{d}{n_k} \log \frac{en_k}{d}}. \tag{A20}$$

Theorem B.2 (Source-Target Generalization Bound of IDEAL)**.** *Let the conditions of Theorem B.1 and Assumptions B.1 to B.4 hold. Then, for any $\beta \in \Delta_K$, with probability at least $1-\delta$, the following holds for any $\Phi \in \mathcal{H}$,*

$$\epsilon_T(\Phi) \le \hat{\epsilon}_\beta(\Phi) + \mathfrak{B}_{src}(\beta, n, M, d, \delta) + \mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta) + \Omega_\beta, \tag{A21}$$

where $\hat{\epsilon}_\beta(\Phi)$ as the empirical source risk in Eq. (A2), $\mathfrak{B}_{src}$ as the source-side bound in Eq. (A20), $\mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta)$ as the discrepancy between the target distribution $\mathbb{P}_T$ and the weighted source mixture $\mathbb{P}_\beta$ in Eq. (A18), and Ω_β as the weighted joint optimal risk in Eq. (A19).

Equivalently, by expanding $\mathfrak{B}_{src}$, we can further get:

$$\epsilon_T(\Phi) \le \hat{\epsilon}_\beta(\Phi) + \sqrt{\frac{1}{2}\left(\sum_{k=1}^{K} \frac{\beta_k^2}{n_k}\right) \log \frac{M}{\delta}} + 2\sum_{k=1}^{K} \beta_k \sqrt{\frac{d}{n_k} \log \frac{en_k}{d}} + \mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta) + \Omega_\beta. \tag{A22}$$

Theorem B.2 indicates that the target-domain generalization bound of IDEAL is associated with the empirical weighted source risk $\hat{\epsilon}_\beta(\Phi)$, the source-side estimation bound $\mathfrak{B}_{src}$, the discrepancy between the target distribution and weighted source mixture, and the weighted joint optimal risk. Theorem B.2 further reveals a capacity trade-off controlled by the number M of intrinsic deviation patterns. Increasing M can improve the coverage of diverse intrinsic deviation vectors and potentially reduce the empirical and approximation errors, but it also enlarges the induced loss class and increases the source-side estimation term through $\log M$. Therefore, M should be properly chosen to balance expressiveness and estimation complexity.

Since the target data are unavailable during training, neither the $\mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta)$ nor Ω_β is directly estimable from the source data. To further simplify the target-side generalization bound, we consider a stronger transfer setting under which the discrepancy term and the weighted joint optimal risk can be tightly controlled. The key is that the target episode distribution $\mathbb{P}_T$ is assumed to lie in the weighted convex hull of the source episode distributions, and that the hypothesis class contains $\Phi^* \in \mathcal{H}$ that performs well on both the target domain and the weighted source mixture.

We now impose the following two assumptions:

Assumption B.5 (Source-Mixture Target)**.** *There exists a source weighting vector $\beta \in \Delta_K$ such that the target episode distribution $\mathbb{P}_T$ is exactly the weighted mixture of the source episode distributions:*

$$\mathbb{P}_T = \mathbb{P}_\beta = \sum_{k=1}^{K} \beta_k \mathbb{P}_{S_k}. \tag{A23}$$

Assumption B.6 (Optimal Hypothesis)**.** *There exists a hypothesis $\Phi^* \in \mathcal{H}$ such that:*

$$\epsilon_T(\Phi^*) + \epsilon_\beta(\Phi^*) \leq \Lambda_{\mathbb{P}_T, \mathbb{P}_\beta}, \tag{A24}$$

where $\Lambda_{\mathbb{P}_T, \mathbb{P}_\beta} \geq 0$ means the joint approximation error. The smaller $\Lambda_{\mathbb{P}_T, \mathbb{P}_\beta}$ is, the better aligned the target domain $\mathbb{P}_T$ is with the weighted source mixture $\mathbb{P}_\beta$ under the hypothesis class $\mathcal{H}$.

Corollary B.3 (Simplified Source-Target Generalization Bound for IDEAL)**.** *Let the conditions of Theorem B.2 and Assumptions B.5 to B.6 hold. With probability at least $1-\delta$, for any $\Phi \in \mathcal{H}$,*

$$\epsilon_T(\Phi) \leq \hat{\epsilon}_\beta(\Phi) + \sqrt{\frac{1}{2}\left(\sum_{k=1}^{K} \frac{\beta_k^2}{n_k}\right) \log \frac{M}{\delta}} + 2\sum_{k=1}^{K} \beta_k \sqrt{\frac{d}{n_k} \log \frac{e n_k}{d}} + \Lambda_{\mathbb{P}_T, \mathbb{P}_\beta}. \tag{A25}$$

Corollary B.3 shows that the target-domain error of IDEAL is upper bounded by the empirical weighted source risk, the source-side estimation term, and a joint approximation error $\Lambda_{\mathbb{P}_T, \mathbb{P}_\beta}$. In particular, when $\Lambda_{\mathbb{P}_T, \mathbb{P}_\beta}$ is small, the dominant estimable term becomes the source-side bound derived in Theorem B.1.

Proof of Theorem B.2 and Corollary B.3. For any source-side weighting vector $\beta \in \Delta_K$, define the weighted source mixture distribution as,

$$\mathbb{P}_\beta = \sum_{k=1}^{K} \beta_k \mathbb{P}_{S_k}. \tag{A26}$$

Following the multi-source domain adaptation [48], for any $\Phi \in \mathcal{H}$, we have:

$$\epsilon_T(\Phi) \leq \epsilon_\beta(\Phi) + \mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta) + \Omega_\beta, \tag{A27}$$

where $\mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta)$ as the discrepancy between the target distribution $\mathbb{P}_T$ and the weighted source mixture $\mathbb{P}_\beta$ in Eq. (A18), and Ω_β as the weighted joint optimal risk in Eq. (A19).

Eq. (A27) remains to upper bound the weighted source risk $\epsilon_\beta(\Phi)$. By Theorem B.1, with probability at least $1-\delta$, the following source-side generalization bound holds,

$$\sup_{\Phi \in \mathcal{H}} |\epsilon_\beta(\Phi) - \hat{\epsilon}_\beta(\Phi)| \leq \mathfrak{B}_{src}(\beta, n, M, d, \delta), \tag{A28}$$

where $\mathfrak{B}_{src}(\beta, n, M, d, \delta)$ is defined in Eq. (A20).

Since the above inequality holds over all $\Phi \in \mathcal{H}$, with probability at least $1-\delta$, it follows that,

$$\epsilon_\beta(\Phi) \leq \hat{\epsilon}_\beta(\Phi) + \mathfrak{B}_{src}(\beta, n, M, d, \delta). \tag{A29}$$

Substituting this bound into the multi-source adaptation inequality yields,

$$\epsilon_T(\Phi) \leq \hat{\epsilon}_\beta(\Phi) + \mathfrak{B}_{src}(\beta, n, M, d, \delta) + \mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta) + \Omega_\beta. \tag{A30}$$

Then, by expanding $\mathfrak{B}_{src}(\beta, n, M, d, \delta)$, we have:

$$\epsilon_T(\Phi) \leq \hat{\epsilon}_\beta(\Phi) + \sqrt{\frac{1}{2}\left(\sum_{k=1}^{K} \frac{\beta_k^2}{n_k}\right) \log \frac{M}{\delta}} + 2\sum_{k=1}^{K} \beta_k \sqrt{\frac{d}{n_k} \log \frac{e n_k}{d}} + \mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta) + \Omega_\beta. \tag{A31}$$

We have completed the proof of Theorem B.2.

Next, under the Assumption B.5, $\mathbb{P}_T$ coincides with the weighted source mixture $\mathbb{P}_\beta$,

$$\mathbb{P}_T = \mathbb{P}_\beta = \sum_{k=1}^{K} \beta_k \mathbb{P}_{S_k}. \tag{A32}$$

Therefore, we can get:

$$\mathrm{D}_{\mathcal{H},\ell}(\mathbb{P}_T, \mathbb{P}_\beta) = 0. \tag{A33}$$

Next, by the definition of $\Omega_\beta = \inf_{\Phi\in\mathcal{H}}(\epsilon_T(\Phi) + \epsilon_\beta(\Phi))$ in Eq. (A19), since the Assumption B.6 guarantees the existence of $\Phi^* \in \mathcal{H}$ such that,

$$\epsilon_T(\Phi^*) + \epsilon_\beta(\Phi^*) \leq \Lambda_{\mathbb{P}_T,\mathbb{P}_\beta}, \tag{A34}$$

which follows that $\Omega_\beta \leq \Lambda_{\mathbb{P}_T,\mathbb{P}_\beta}$.

Substituting the above two inequalities into Eq. (A27), we can get:

$$\epsilon_T(\Phi) \leq \epsilon_\beta(\Phi) + \Lambda_{\mathbb{P}_T,\mathbb{P}_\beta}. \tag{A35}$$

On the other hand, by Theorem B.1, with probability at least $1-\delta$, the source-side generalization bound holds for any $\Phi \in \mathcal{H}$,

$$\epsilon_\beta(\Phi) \leq \hat{\epsilon}_\beta(\Phi) + \mathfrak{B}_{src}(\beta, n, M, d, \delta). \tag{A36}$$

Combining the above two inequalities, we have:

$$\epsilon_T(\Phi) \leq \hat{\epsilon}_\beta(\Phi) + \mathfrak{B}_{src}(\beta, n, M, d, \delta) + \Lambda_{\mathbb{P}_T,\mathbb{P}_\beta}. \tag{A37}$$

Equivalently,

$$\epsilon_T(\Phi) \leq \hat{\epsilon}_\beta(\Phi) + \sqrt{\frac{1}{2}\left(\sum_{k=1}^{K}\frac{\beta_k^2}{n_k}\right)\log\frac{M}{\delta}} + 2\sum_{k=1}^{K}\beta_k\sqrt{\frac{d}{n_k}\log\frac{en_k}{d}} + \Lambda_{\mathbb{P}_T,\mathbb{P}_\beta}, \tag{A38}$$

where $\mathfrak{B}_{src}$ means the source-side generalization bound in Eq. (A20), and $\Lambda_{\mathbb{P}_T,\mathbb{P}_\beta}$ means the joint approximation error between the target episode distribution and the weighted source mixture.

A particularly clean special case arises when the hypothesis class contains a predictor with zero joint error on the target domain and the weighted source mixture.

In this case, we have $\Lambda_{\mathbb{P}_T,\mathbb{P}_\beta} = 0$, for any $\Phi \in \mathcal{H}$, the Corollary B.3 reduces to,

$$\epsilon_T(\Phi) \leq \hat{\epsilon}_\beta(\Phi) + \sqrt{\frac{1}{2}\left(\sum_{k=1}^{K}\frac{\beta_k^2}{n_k}\right)\log\frac{M}{\delta}} + 2\sum_{k=1}^{K}\beta_k\sqrt{\frac{d}{n_k}\log\frac{en_k}{d}}, \tag{A39}$$

where the source-target error is completely reduced to the empirical weighted source risk plus the source-side estimation term. Therefore, under this ideal setting, Eq. (A39) bounds the source-side term sufficiently to control the target-domain generalization performance. □

C Experimental Details

In this section, we first describe the eight anomaly detection datasets utilized in our experiments (Sec. C.1), followed by the competing baselines under cross-dataset and in-dataset evaluation protocols (Sec. C.2). We then detail the episode construction strategy for discriminative FSAD during training and inference (Sec. C.3). Finally, we provide implementation details of IDEAL (Sec. C.4).

C.1 Anomaly Detection Datasets

We conduct extensive experiments on eight real-world Anomaly Detection (AD) datasets, including five industrial defect inspection datasets: MVTecAD [8], VisA [94], AITEX [62], MPDD [33], BTAD [52]; and three medical image datasets: BraTS [50] (for brain tumor segmentation), Liver [4] (for liver tumor segmentation), RESC [28] (for retinal edema segmentation). We then describe each anomaly detection dataset in turn below.

MVTecAD [8] is a widely used benchmark for evaluating anomaly detection methods in industrial inspection applications. This dataset includes over 5000 high-resolution images spanning 15 object and texture categories. For each category, the training set consists exclusively of anomaly-free images, while the test set includes both defective and defect-free samples.

VisA [94] is a large-scale industrial anomaly detection dataset comprising 10821 high-resolution color images, including 9621 normal samples and 1200 anomalous samples. This dataset covers 12 object categories across 3 domains and provides both image-level and pixel-level annotations. The anomalous samples exhibit diverse defect types, including surface defects such as scratches, dents, color spots, and cracks, as well as structural defects such as misplacement and missing parts.

AITEX [62] is a textile fabric anomaly detection dataset comprising 245 images from 7 fabric types, including 140 defect-free images (20 images for each fabric type) and 105 images with various defects. Following the evaluation protocol, we use its test set for evaluation.

MPDD [33] is a metal-part defect detection dataset designed to benchmark visual anomaly detection methods under complex industrial conditions, which contains over 1000 images of metal parts with pixel-precise defect annotation masks. It contains 1346 images across 6 metal-part categories, with 888 normal images for training and 458 test images (176 normal and 282 anomalous samples).

BTAD [52] is a industrial anomaly detection dataset introduced in VT-ADL. It contains 2830 images from 3 industrial products, covering both body and surface defects. The training set consists of 1799 normal images, while the test set contains 451 normal images and 290 anomalous images.

BraTS [50] is a large-scale medical image segmentation dataset for multimodal brain tumor analysis. It contains 2000 glioma cases, corresponding to 8000 mpMRI scans across four modalities: T1, T1Gd, T2, and FLAIR. Each case provides expert-annotated tumor sub-regions, enabling evaluation of abnormal tissue detection and localization in medical imaging scenarios.

Liver [4] is a medical anomaly detection dataset reorganized from BMAD for liver tumor detection and localization. The dataset contains 3201 axial CT slices, including 1542 training slices, 1493 test slices, and 166 validation slices, with a spatial resolution of 512×512. Pixel-level segmentation masks are provided for anomaly localization evaluation.

RESC [28] is a retinal OCT dataset originally developed for retinal edema segmentation. Following BMAD, we use its reorganized anomaly detection and localization split, where retinal edema regions are treated as anomalies. It contains 6217 OCT images with a spatial resolution of 512×1024, including 4297 training images, 1805 test images, and 115 validation images. Pixel-level segmentation masks are provided for edema regions, enabling evaluation of anomaly detection.

C.2 Competing Baselines

We compare IDEAL with two groups of closely related methods: (1) *Generalist Methods*, *i.e.*, detectors are training-free or trained once on auxiliary data: based on normal-only references (InCTRL [93], ResAD [77], WinCLIP [32], FoundAD [84]), and normal and anomalous references (NAGL [73]); (2) *Specalist Methods*, *i.e.*, detectors require dataset-specific training: based on normal-only references (RegAD [29], PromptAD [36]), and based on full-shot normal data and abnormal-only references: DRA [18] and AHL [91].

The crucial configurations of different AD baselines as follows. 1) **WinCLIP** [32]: We follow the original setting with the LAION-400M pre-trained CLIP ViT-B/16+ backbone. The dense window features are extracted with stride 1 over ViT patch embeddings, and multi-scale predictions from 2×2 patch windows, 3×3 patch windows, and the image-level token are aggregated by harmonic averaging. 2) **InCTRL** [93]: We use the pre-trained ViT-B/16+ as backbone with the initial learning rate as $1e^{-3}$ by default. The prompt ensemble contains 231 text prompts per object category, including 147 normal prompts and 84 anomalous prompts. The training epochs is set to 10 with batch size of 4. 3) **ResAD** [77]: The number of the NF model's layers is set to 8. The total training epochs are set as 100 with a batch size of 16. The learning rate is set to $1e^{-5}$ initially and dropped by 0.1 after 75 epochs. 4) **FoundAD** [84]: The pre-trained DINOv3 ViT-B is adopted as the frozen visual encoder. All the training and testing images are resized to 512×512. The nonlinear manifold projector is configured with a prediction depth of 6 and an embedding dimension of 384. The model is trained for 100 epochs with a batch size of 8, an initial learning rate of $1e^{-3}$, and a weight decay of $1e^{-4}$. The learning rate is kept constant by default. 5) **NAGL** [73]: We follow the official implementation

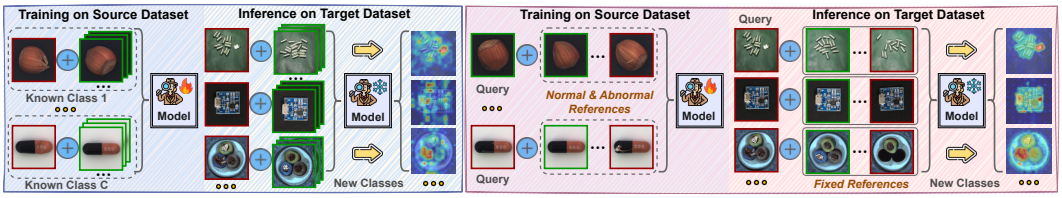

Figure A1: Illustration of **(a)** generalist FSAD setting (*e.g.*, used in InCTRL [93]); **(b)** discriminative FSAD setting with few-shot normal and abnormal references, our model effectively leverages the discriminative clues of abnormal references relative to normal references to enhance detection.

and use the pre-trained DINOv2 ViT-S/14 as the frozen visual backbone. All the training and testing images are resized to 448×448. We tune the number of learnable proxies $\{10, 25, 30, 40\}$ and set it as 25. The model is trained for 20 epochs with a batch size of 16. The initial learning rate is set to $1e^{-5}$ and optimized by AdamW. The learning rate is decayed by 0.1 at the 10-th and 15-th epochs. 6) **RegAD** [29]: We use ResNet-18 as the backbone, followed by a convolution-based encoder and predictor. The input images are resized to 224×224. The model is trained for 50 epochs with a batch size of 16 using momentum SGD. The initial learning rate is set to $1e^{-4}$ and decayed by a single-cycle cosine schedule. 7) **PromptAD** [36]: We follow the official implementation and use ViT-B/16+ as the visual backbone. The input resolution is set to 240 for the shorter edge after bicubic resizing. The loss weight λ of PromptAD is set to 0.001. The prompt learning optimizer uses an initial learning rate of 0.002, momentum of 0.9, and weight decay of $5e^{-4}$. The 3-rd and 8-th layer features are used for visual feature memory in PromptAD. 8) **DRA** [18]: We follow the official implementation and use ImageNet pre-trained ResNet-18 as the backbone. The extracted 512-dimensional feature map is fed into the abnormality and normality learning heads. The patch-wise classifier is implemented by a 1×1 convolutional layer. The model is trained for 30 epochs with 20 iterations per epoch and a batch size of 48. Adam is used for optimization with an initial learning rate of $1e^{-3}$ and a weight decay of $1e^{-2}$. 9) **AHL** [91]: We follow the official setting and use DRA as the base OSAD detector. Normal samples are first partitioned into 3 clusters by K-means, and $T = 6$ anomaly distribution subsets are generated by combining randomly selected normal clusters with labeled anomaly examples. Adam is used as the optimizer, and the learning rate is set to $2e^{-4}$ for heterogeneous base models and $2e^{-3}$ for the unified anomaly detection model. The score estimator is implemented with a two-layer LSTM with hidden dimension 6, followed by a fully connected layer with 12 hidden nodes.

C.3 Episode Preparation

The proposed IDEAL is trained and evaluated in an episodic manner, as shown in Fig. A1. In each episode, a limited reference set containing normal and anomalous reference samples is provided, and the model is required to detect anomalies in the corresponding query samples.

Training Episode. IDEAL is trained on the auxiliary dataset, where each training episode is constructed to simulate the few-shot inference scenario, consisting of a query image together with a few normal and anomalous reference images sampled from the auxiliary dataset. Specifically, we first randomly select 600 samples by default from the entire test set as training queries, including both normal and abnormal samples. Based on the query, we randomly sample normal reference images from the corresponding training set and anomalous reference images from the corresponding test set, where all anomalous references are restricted to a single anomaly type. During training, each episode is constructed by grouping a query with its corresponding normal and anomalous reference samples, and the model is optimized episode by episode.

Inference Episode. IDEAL is evaluated on the target dataset without further retraining/tuning. For each target dataset, we pre-construct a fixed reference set containing normal and anomalous reference images, using the same sampling strategy as in training. These references are kept fixed throughout the evaluation. When an arbitrary query image from the target dataset, it is combined with the corresponding fixed normal and anomalous references to form an inference episode, which is then fed into our model for anomaly detection.

C.4 Implementation Details

By default, we adopt the public pre-trained ViT/S-14 as the feature extractor in our experiments. We freeze the parameters of the feature extractor and only update the parameters of the proposed IDE component. The IDE component is implemented as a cross-attention block followed by a feed-forward MLP block. The cross-attention block has an embedding dimension of 384, 12 attention heads, an attention dropout rate of 0.1, and sinusoidal positional encoding. The MLP consists of two linear layers with input/output dimension 384 and hidden dimension 1536, with GELU used as the activation function. For the learnable vectors $\mathcal{T}$ in Eq. (5), we instantiate them as a learnable embedding table with M (default by 45) vectors, each with a dimension of 384. During training, all the training and testing images are resized to 448×448 resolution for injection into the feature extractor. AdamW is used as the optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The weight decay is set to $1e^{-4}$, and AMSGrad is enabled for more stable optimization. The initial learning rate is set to 0.001 with a warm-up cosine scheduler, where the learning rate is first linearly increased from $1e^{-5}$ to 0.001 during the first 2 epochs and then cosine-decayed to 1% of the base learning rate. The training epochs of IDEAL is set to 20 with a batch size of 16 on a single NVIDIA GeForce RTX 3090 GPU. By default, the number k of normal neighbors in Eq. (2) is set to 12, the number r of retained principal components is set to 4, α in Eq. (4) is set to 0.8, and the number of intrinsic deviation vectors M in Eq. (6) is set to 45. The loss weights λ_1 and λ_2 in Eq. (7) are set to 1.0 and 0.8, respectively. More hyper-parameter ablations, please see Appendix E.

D NVE as a Plug-and-Play Component

The purpose of NVE (Sec. 4.1) is to isolate anomaly-relevant deviation representations by eliminating nuisance normal variations that may lead to noisy deviations from normality. Since our NVE component operates at the feature level, it can be readily incorporated with existing residual-based FSAD methods without modifying their core architectures. Tab. A1 presents image-level and pixel-level AUROCs of four residual-based FSAD methods (ResAD [77], InCTRL [93], RegAD [29], NAGL [73]) after replacing their original residual features with NVE-denoised residual features. We can observe that our NVE component consistently improves the performance of these methods in MVTecAD and VisA datasets. This confirms that nuisance normal variations are indeed present in the residual features used by these methods, and that NVE suppresses such variations to produce cleaner residual features.

Table A1: **Enabling the existing residual-based FSAD methods**, in which their residual features are replaced with NVE-denoised residual features.

METHODS		MVTECAD	VISA
N1	InCTRL	(93.3 $_{\uparrow 0.8}$ / 94.7 $_{\uparrow 1.3}$)	(83.4 $_{\uparrow 1.2}$ / 93.5 $_{\uparrow 0.4}$)
	ResAD	(83.5 $_{\uparrow 2.8}$ / 86.0 $_{\uparrow 1.7}$)	(79.2 $_{\uparrow 2.0}$ / 90.0 $_{\uparrow 1.7}$)
	RegAD	(93.1 $_{\uparrow 1.5}$ / 93.8 $_{\uparrow 0.7}$)	(85.2 $_{\uparrow 1.9}$ / 92.0 $_{\uparrow 0.4}$)
N1A1	NAGL	(95.3 $_{\uparrow 0.2}$ / 96.6 $_{\uparrow 0.5}$)	(89.4 $_{\uparrow 0.9}$ / 97.6 $_{\uparrow 0.1}$)
N4	InCTRL	(94.7 $_{\uparrow 0.4}$ / 94.5 $_{\uparrow 0.5}$)	(90.2 $_{\uparrow 0.7}$ / 94.5 $_{\uparrow 0.3}$)
	ResAD	(89.1 $_{\uparrow 2.0}$ / 88.1 $_{\uparrow 1.2}$)	(84.7 $_{\uparrow 1.6}$ / 91.0 $_{\uparrow 0.9}$)
	RegAD	(94.9 $_{\uparrow 0.9}$ / 95.5 $_{\uparrow 0.4}$)	(89.9 $_{\uparrow 0.4}$ / 94.6 $_{\uparrow 1.3}$)
N4A1	NAGL	(97.0 $_{\uparrow 0.1}$ / 97.3 $_{\uparrow 0.3}$)	(91.8 $_{\uparrow 0.6}$ / 97.8 $_{\uparrow 0.1}$)
N8	InCTRL	(96.1 $_{\uparrow 0.3}$ / 95.0 $_{\uparrow 0.6}$)	(90.7 $_{\uparrow 0.9}$ / 95.5 $_{\uparrow 0.2}$)
	ResAD	(90.5 $_{\uparrow 1.1}$ / 88.5 $_{\uparrow 0.7}$)	(85.0 $_{\uparrow 1.3}$ / 91.2 $_{\uparrow 0.3}$)
	RegAD	(95.3 $_{\uparrow 0.3}$ / 96.1 $_{\uparrow 0.3}$)	(92.5 $_{\uparrow 0.6}$ / 95.1 $_{\uparrow 0.4}$)
N8A4	NAGL	(97.4 $_{\uparrow 0.2}$ / 97.5 $_{\uparrow 0.4}$)	(92.0 $_{\uparrow 0.5}$ / 98.1 $_{\uparrow 0.2}$)

E Hyper-parameter Ablation

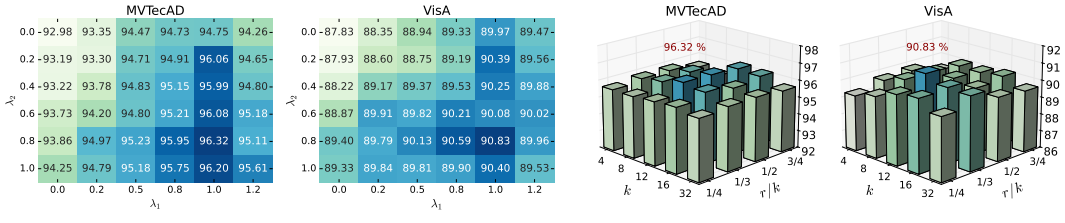

Figure A2: Under the N1A1 setting. **Left**: hyper-parameter study with different λ_1 and λ_2 in Eq. (7) on MVTecAD and VisA. **Right**: hyper-parameter study with the number k of normal neighbors in Eq. (2) and the number r of retained principal components in the process of NVE (Sec. 4.1).

Table A2: Under the `N1A1` setting. **Left**: hyper-parameter study with α in Eq. (4) on MVTecAD and VisA. **Right**: the sensitivity among the loss terms in Eq. (10).

α in Eq. (4)	MVTecAD	VisA
$\alpha = 0.2$	(95.5 / 96.0)	(89.8 / 97.1)
$\alpha = 0.5$	(**96.2** / **96.5**)	(**90.5** / **97.6**)
$\alpha = 0.8$	(**96.3** / **96.8**)	(**90.8** / **97.8**)
$\alpha = 1.0$	(96.0 / 96.2)	(90.3 / 97.3)

$\mathcal{L}_{Focal}$	$\mathcal{L}_{Dice}$	$\mathcal{L}_{BCE}$	$\mathcal{L}_{Dual}$	MVTecAD	VisA
✓	✓			(93.8 / 95.9)	(86.7 / 95.6)
✓	✓	✓		(94.3 / 96.3)	(88.2 / **97.2**)
✓	✓		✓	(**94.8** / **96.6**)	(**88.7** / **97.2**)
✓	✓	✓	✓	(**96.3** / **96.8**)	(**90.8** / **97.8**)

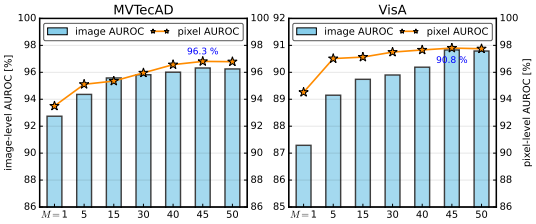

Figure A3: Under the `N1A1` setting. Ablation study with the number M of intrinsic deviation vectors in Eq. (6).

We first tune three key arguments in IDEAL: loss weights λ_1 and λ_2 in Eq. (7), the number M of intrinsic deviation vectors in Eq. (6). Fig. A2 (**Left**) reports the image-level AUROC with different λ_1 and λ_2 on the MVTecAD and VisA datasets. We can observe that the performance improves gradually as λ_1 and λ_2 increase, indicating that both the discriminability and orthogonality constraints are beneficial for learning the intrinsic deviation vectors. However, an excessively large λ_2 becomes detrimental to training, because the orthogonality term may force the intrinsic deviation vectors apart even when they share correlated patterns. As shown in Fig. A3, we investigate the number M of intrinsic deviation vectors in Eq. (6) for a wide range, and present image-level and pixel-level AUROCs on the MVTecAD and VisA datasets. We observe that increasing M consistently improves the performance of IDEAL, with the improvement becoming marginal at $M = 45$. This suggests that more intrinsic deviation vectors enable better capture of discriminative orthogonal deviation directions, while the benefit tends to saturate once sufficient capacity is reached. Moreover, we also study the effect of the number k of normal neighbors in Eq. (2) and the number r of retained principal components in the process of NVE (Sec. 4.1). Fig. A2 (**Right**) shows that IDEAL is generally robust to both k and r. A moderate number k of normal neighbors yields better performance, as it provides a more reliable estimate of local normal variations, whereas an overly large k may introduce less relevant neighbors. Similarly, retaining an appropriate proportion of principal components r helps preserve dominant normal patterns while filtering out noisy directions. Tab. A2 (**Left**) shows that IDEAL is robust to the choice of $\alpha \in [0.5, 0.8]$ in Eq. (4), and $\alpha = 0.8$ yields the best overall performance, suggesting that a moderate elimination strength provides a favorable balance. Empirically, $\lambda_1 = 1.0$, $\lambda_2 = 0.8$, $M = 45$, $\alpha = 0.8$, $k = 12$, $r = 4$ is a recommended configuration.

F Ablation Study on Loss Terms

We conduct an ablation study among the loss terms of Eq. (10) in Tab. A2 (**Right**). We observe that $\mathcal{L}_{Focal}$ and $\mathcal{L}_{Dice}$ provide the basic pixel-level supervision for anomaly localization, while $\mathcal{L}_{BCE}$ improves image-level anomaly discrimination. Moreover, adding the $\mathcal{L}_{Dual}$ consistently improves performance, suggesting that encouraging the learning of intrinsic deviation vectors helps extract more discriminative deviations from normality, thereby enhancing detection. Combining all loss terms yields the best results on both MVTecAD and VisA, confirming their complementary roles.

G Limitations and Social Impacts

The proposed IDEAL is currently evaluated only on image anomaly detection datasets, including industrial and medical images. It would be highly valuable to further apply our framework to other data modalities, such as tabular and time-series data, to more comprehensively assess the generalizability of IDEAL. This work does not introduce specific ethical concerns or social impacts beyond those commonly associated with anomaly detection. All datasets used in our experiments are publicly available, and no private user data or personally identifiable information is collected.

Table A3: **Comparison results on MVTecAD** under various few-shot AD settings. Best results are in **red bold**, with second best results in **blue bold**.

SETUPS	METHODS	IMAGE-LEVEL			PIXEL-LEVEL		
		AUROC	AP	F1-Max	AUROC	PRO	F1-Max
N1	WinCLIP [32]	93.1	96.5	93.6	93.4	83.4	42.2
	InCTRL [93]	92.5	96.2	93.7	93.4	83.3	45.9
	ResAD [77]	80.7	85.8	86.8	84.3	70.6	32.0
	FoundAD [84]	89.9	91.0	83.6	90.7	88.7	56.8
	RegAD [29]	91.6	90.8	90.9	93.1	87.3	54.4
	PromptAD [36]	92.8	92.7	91.3	93.4	90.9	56.6
A1	DRA [18]	92.2	96.1	93.1	92.9	91.2	51.7
	AHL [91]	91.7	94.3	93.3	93.6	90.8	54.2
N1A1	NAGL [73]	**95.1**	**96.9**	**95.4**	**96.1**	**92.9**	**58.7**
	IDEAL (ours)	**96.3**	**97.3**	**95.7**	**96.8**	**93.4**	**59.1**
N2	WinCLIP [32]	93.4	96.5	94.1	93.6	84.2	42.1
	InCTRL [93]	92.7	96.5	93.8	93.7	83.7	46.6
	ResAD [77]	85.7	89.0	88.2	86.3	72.6	34.2
	FoundAD [84]	90.3	91.8	94.1	92.8	89.9	58.4
	RegAD [29]	92.6	93.2	91.8	93.5	91.8	56.5
	PromptAD [36]	93.5	92.9	91.5	93.8	91.3	57.2
N2A1	NAGL [73]	**96.0**	**97.8**	**96.4**	**96.6**	**93.3**	**59.8**
	IDEAL (ours)	**97.4**	**98.8**	**96.9**	**97.5**	**93.5**	**60.0**
N4	WinCLIP [32]	93.9	97.3	95.1	93.8	84.5	43.1
	InCTRL [93]	94.3	96.8	95.0	94.0	83.8	46.9
	ResAD [77]	87.1	89.4	88.7	86.9	73.1	36.2
	FoundAD [84]	92.9	92.9	95.6	93.2	90.7	**60.5**
	RegAD [29]	94.0	94.9	93.1	95.1	92.3	60.1
	PromptAD [36]	94.9	93.0	91.7	94.6	92.4	57.5
N4A1	NAGL [73]	**96.9**	**97.5**	**96.6**	**97.0**	**93.4**	59.9
	IDEAL (ours)	**98.2**	**99.0**	**97.6**	**97.5**	**93.9**	**62.1**
N8	WinCLIP [32]	95.5	97.8	95.1	94.2	85.9	45.6
	InCTRL [93]	95.8	97.8	95.5	94.4	84.7	48.1
	ResAD [77]	89.4	90.2	89.5	87.8	74.2	37.8
	FoundAD [84]	93.2	93.9	96.7	93.5	91.4	60.3
	RegAD [29]	95.0	96.2	94.2	95.8	**93.8**	**61.0**
	PromptAD [36]	96.0	93.3	91.8	94.8	92.9	57.6
A4	DRA [18]	93.7	96.0	93.3	94.0	93.3	55.5
	AHL [91]	92.9	94.7	94.0	94.5	93.5	57.4
N8A4	NAGL [73]	**97.2**	**98.7**	**97.3**	**97.1**	**93.8**	60.7
	IDEAL (ours)	**98.6**	**99.2**	**98.0**	**97.8**	**94.2**	**62.4**

H Additional Results

We provide complete results on different datasets, including MVTecAD (Tab. A3), VisA (Tab. A4), AITEX (Tab. A5), MPDD (Tab. A6), BTAD (Tab. A7), BraTS (Tab. A8), Liver (Tab. A9), and RESC (Tab. A10). The details of each dataset, please see the Appendix C.1.

Table A4: **Comparison results on VisA** under various few-shot AD settings. Best results are in **red bold**, with second best results in **blue bold**.

SETUPS	METHODS	IMAGE-LEVEL			PIXEL-LEVEL		
		AUROC	AP	F1-Max	AUROC	PRO	F1-Max
N1	WinCLIP [32]	78.7	80.3	80.2	94.7	79.2	32.9
	InCTRL [93]	82.2	84.1	81.6	93.1	65.2	35.2
	ResAD [77]	77.2	72.1	76.2	88.3	70.9	32.1
	FoundAD [84]	85.3	**89.9**	82.8	96.2	90.5	40.6
	RegAD [29]	83.3	82.5	85.6	91.6	89.8	32.7
	PromptAD [36]	83.8	87.8	**86.1**	89.5	87.8	**41.7**
A1	DRA [18]	82.3	84.6	80.7	90.3	77.3	40.5
	AHL [91]	82.0	85.5	81.4	91.9	81.6	40.8
N1A1	NAGL [73]	**88.5**	89.2	85.2	**97.5**	**91.0**	41.5
	IDEAL (ours)	**90.8**	**91.3**	**87.2**	**97.8**	**92.1**	**42.1**
N2	WinCLIP [32]	84.3	85.7	82.7	94.8	80.1	33.1
	InCTRL [93]	85.6	89.0	85.8	93.7	74.6	36.3
	ResAD [77]	82.4	73.5	77.5	89.8	71.9	32.7
	FoundAD [84]	88.5	90.6	83.2	96.5	90.9	40.9
	RegAD [29]	87.4	88.4	86.0	92.7	90.4	38.1
	PromptAD [36]	86.6	89.7	**87.5**	89.8	89.1	42.1
N2A1	NAGL [73]	**89.1**	**90.7**	86.0	**97.5**	**91.5**	**43.2**
	IDEAL (ours)	**91.3**	**91.9**	**87.9**	**97.9**	**92.9**	**44.5**
N4	WinCLIP [32]	84.5	85.8	83.2	95.1	80.0	34.9
	InCTRL [93]	89.5	89.7	86.5	94.2	74.3	36.9
	ResAD [77]	83.1	75.4	78.9	90.1	72.2	36.4
	FoundAD [84]	91.0	90.9	84.6	96.7	90.4	42.3
	RegAD [29]	89.5	90.8	87.1	93.3	**91.5**	42.5
	PromptAD [36]	87.1	90.1	**87.9**	90.6	89.5	42.8
N4A1	NAGL [73]	**91.2**	**91.7**	87.1	**97.7**	91.3	**44.0**
	IDEAL (ours)	**92.7**	**92.6**	**89.1**	**98.0**	**93.5**	**44.8**
N8	WinCLIP [32]	86.1	87.5	83.7	95.1	81.5	35.3
	InCTRL [93]	89.8	90.9	86.6	95.3	74.9	37.8
	ResAD [77]	83.7	77.4	79.5	90.9	73.9	38.3
	FoundAD [84]	91.3	91.1	84.8	96.8	90.4	43.7
	RegAD [29]	**91.9**	92.0	88.3	94.7	92.0	43.4
	PromptAD [36]	87.9	90.7	88.2	92.2	90.6	43.5
A4	DRA [18]	86.8	91.8	87.8	90.7	87.9	42.4
	AHL [91]	85.9	90.6	88.3	92.5	87.0	41.9
N8A4	NAGL [73]	91.5	**92.8**	**88.5**	**97.9**	**92.1**	**45.3**
	IDEAL (ours)	**93.1**	**92.9**	**89.4**	**98.3**	**93.6**	**45.9**

Table A5: **Comparison results on AITEX** under various few-shot AD settings. Best results are in **red bold**, with second best results in **blue bold**.

SETUPS	METHODS	IMAGE-LEVEL			PIXEL-LEVEL		
		AUROC	AP	F1-Max	AUROC	PRO	F1-Max
N1	WinCLIP [32]	73.1	**48.4**	51.1	78.9	58.3	**22.1**
	InCTRL [93]	**74.9**	41.4	60.3	78.3	**60.5**	11.7
	ResAD [77]	70.5	37.4	51.6	72.7	51.9	10.3
	FoundAD [84]	71.4	42.0	54.6	80.4	57.1	10.5
	RegAD [29]	72.0	47.1	58.2	80.2	56.6	11.4
	PromptAD [36]	71.9	47.8	**60.5**	**81.4**	58.2	15.8
A1	DRA [18]	72.3	45.1	55.0	75.3	54.9	15.7
	AHL [91]	70.7	39.8	53.3	76.3	53.6	21.7
N1A1	NAGL [73]	71.6	47.3	55.3	75.5	58.4	14.3
	IDEAL (ours)	**75.8**	**50.2**	**61.2**	**82.7**	**62.0**	**25.5**
N2	WinCLIP [32]	73.5	**52.2**	52.6	79.3	58.6	**22.9**
	InCTRL [93]	**75.2**	41.8	60.4	78.7	**60.4**	12.9
	ResAD [77]	75.0	47.2	59.3	75.1	55.7	13.4
	FoundAD [84]	72.3	43.1	56.5	81.4	59.3	12.1
	RegAD [29]	73.5	51.1	59.9	**82.0**	59.1	14.7
	PromptAD [36]	73.8	48.1	**60.6**	81.8	58.8	16.9
N2A1	NAGL [73]	73.5	48.3	58.2	79.8	58.7	15.9
	IDEAL (ours)	**77.6**	**53.1**	**61.7**	**85.3**	**62.6**	**26.9**
N4	WinCLIP [32]	73.8	52.5	52.9	79.5	60.7	**23.2**
	InCTRL [93]	**75.6**	43.7	60.2	80.4	**64.5**	13.7
	ResAD [77]	75.2	50.6	59.4	76.7	57.1	14.1
	FoundAD [84]	73.2	47.0	58.2	82.8	63.3	14.7
	RegAD [29]	74.6	**54.0**	60.1	**83.7**	63.3	15.5
	PromptAD [36]	74.1	50.8	**61.4**	83.4	62.9	18.6
N4A1	NAGL [73]	**75.6**	50.2	59.5	80.9	60.4	19.0
	IDEAL (ours)	**79.5**	**58.8**	**63.4**	**86.1**	**66.1**	**29.5**
N8	WinCLIP [32]	74.8	55.0	53.4	80.5	62.6	28.3
	InCTRL [93]	75.9	44.1	60.6	80.6	65.2	14.6
	ResAD [77]	75.4	53.8	59.4	79.3	57.4	15.8
	FoundAD [84]	74.3	52.1	**64.6**	83.0	64.1	15.7
	RegAD [29]	75.8	56.5	64.2	83.9	**66.4**	17.5
	PromptAD [36]	74.4	56.0	61.9	**84.7**	64.3	20.2
A4	DRA [18]	75.7	**58.3**	56.5	77.8	56.5	23.2
	AHL [91]	75.1	56.3	52.6	78.8	52.6	**29.8**
N8A4	NAGL [73]	**80.5**	54.8	60.5	82.3	65.2	20.5
	IDEAL (ours)	**83.9**	**61.1**	**68.2**	**86.8**	**73.0**	**30.3**

Table A6: **Comparison results on MPDD** under various few-shot AD settings. Best results are in **red bold**, with second best results in **blue bold**.

SETUPS	METHODS	IMAGE-LEVEL			PIXEL-LEVEL		
		AUROC	AP	F1-Max	AUROC	PRO	F1-Max
N1	WinCLIP [32]	70.2	70.6	80.7	94.3	88.1	31.3
	InCTRL [93]	74.8	76.3	80.4	93.1	84.5	32.3
	ResAD [77]	67.7	67.5	75.0	92.7	84.1	29.8
	FoundAD [84]	70.6	70.1	84.5	95.4	92.5	37.9
	RegAD [29]	69.7	70.6	83.6	92.5	85.7	35.5
	PromptAD [36]	75.8	76.2	83.1	95.1	91.4	**39.5**
A1	DRA [18]	73.4	75.3	82.3	93.9	92.1	36.7
	AHL [91]	73.3	76.8	82.7	94.3	92.7	39.1
N1A1	NAGL [73]	**77.1**	**79.3**	**84.6**	**96.5**	**92.8**	39.0
	IDEAL (ours)	**78.5**	**80.5**	**86.8**	**97.9**	**94.1**	**40.6**
N2	WinCLIP [32]	70.8	74.1	82.5	94.7	88.8	32.4
	InCTRL [93]	75.9	78.6	83.4	93.5	86.3	34.4
	ResAD [77]	68.4	69.6	76.1	94.4	86.8	31.7
	FoundAD [84]	73.9	80.9	84.6	95.7	92.6	42.7
	RegAD [29]	71.9	78.1	**86.3**	94.5	86.1	39.2
	PromptAD [36]	76.8	76.9	83.8	96.1	91.5	41.6
N2A1	NAGL [73]	**80.2**	**82.8**	86.1	**97.3**	**93.5**	**43.3**
	IDEAL (ours)	**81.6**	**84.5**	**88.5**	**98.0**	**94.7**	**44.5**
N4	WinCLIP [32]	71.0	75.1	83.0	95.4	89.3	33.4
	InCTRL [93]	79.4	79.1	83.7	93.7	89.0	35.2
	ResAD [77]	70.3	72.2	77.9	95.7	86.9	33.3
	FoundAD [84]	77.6	81.2	85.4	95.8	92.8	43.9
	RegAD [29]	76.3	78.6	**87.2**	95.2	86.3	42.3
	PromptAD [36]	78.3	80.9	85.1	96.5	91.6	**45.6**
N4A1	NAGL [73]	**81.4**	**84.3**	86.7	**97.7**	**93.9**	44.6
	IDEAL (ours)	**87.6**	**89.4**	**89.1**	**98.5**	**95.7**	**46.7**
N8	WinCLIP [32]	74.7	75.5	83.3	95.6	89.2	34.3
	InCTRL [93]	**83.2**	84.5	85.5	94.1	89.7	35.6
	ResAD [77]	75.1	75.7	81.5	96.1	88.1	34.7
	FoundAD [84]	80.3	83.3	86.5	96.2	92.9	**47.3**
	RegAD [29]	80.8	79.1	87.7	95.7	88.4	43.9
	PromptAD [36]	80.6	86.2	86.4	96.5	92.3	45.9
A4	DRA [18]	81.8	83.9	82.0	94.9	92.5	44.7
	AHL [91]	81.0	84.5	83.8	95.2	93.7	46.3
N8A4	NAGL [73]	83.1	**86.5**	**88.6**	**97.9**	**94.1**	45.7
	IDEAL (ours)	**88.0**	**90.4**	**91.5**	**98.7**	**95.9**	**48.7**

Table A7: **Comparison results on BTAD** under various few-shot AD settings. Best results are in **red bold**, with second best results in **blue bold**.

SETUPS	METHODS	IMAGE-LEVEL			PIXEL-LEVEL		
		AUROC	AP	F1-Max	AUROC	PRO	F1-Max
N1	WinCLIP [32]	84.5	88.1	75.7	94.3	66.5	46.9
	InCTRL [93]	90.5	91.1	86.5	92.9	67.8	54.7
	ResAD [77]	78.4	80.3	71.7	92.3	72.9	38.7
	FoundAD [84]	91.4	91.9	87.9	91.9	73.9	51.8
	RegAD [29]	86.6	87.8	85.5	94.5	72.8	45.7
	PromptAD [36]	89.6	90.8	87.7	92.4	72.5	59.5
A1	DRA [18]	90.5	91.7	91.3	93.5	72.3	53.4
	AHL [91]	90.9	92.3	91.9	93.8	73.2	57.8
N1A1	NAGL [73]	**92.0**	**95.6**	**92.2**	**95.8**	**75.5**	**60.6**
	IDEAL (ours)	**93.4**	**96.0**	**93.1**	**97.0**	**76.9**	**62.3**
N2	WinCLIP [32]	85.0	88.9	76.4	95.5	67.1	48.4
	InCTRL [93]	93.4	92.9	88.1	93.0	69.2	56.1
	ResAD [77]	80.1	81.4	72.6	92.9	73.0	41.4
	FoundAD [84]	**94.0**	92.9	88.5	93.2	75.7	53.6
	RegAD [29]	90.8	89.5	86.9	95.6	73.4	46.4
	PromptAD [36]	90.3	91.5	88.5	92.6	72.8	60.5
N2A1	NAGL [73]	93.5	**96.4**	**92.5**	**96.3**	**76.7**	**62.2**
	IDEAL (ours)	**95.6**	**97.2**	**93.2**	**97.2**	**77.0**	**63.7**
N4	WinCLIP [32]	87.4	90.4	77.4	95.9	69.5	49.0
	InCTRL [93]	93.9	93.5	89.1	93.6	72.6	57.8
	ResAD [77]	81.5	82.1	74.9	95.1	76.5	44.5
	FoundAD [84]	**94.3**	93.4	89.7	94.0	**77.4**	60.4
	RegAD [29]	91.7	91.2	90.6	96.3	75.8	51.7
	PromptAD [36]	90.8	92.3	89.7	93.2	74.5	61.3
N4A1	NAGL [73]	93.7	**96.7**	**93.3**	**96.5**	76.5	**62.3**
	IDEAL (ours)	**95.9**	**97.8**	**93.5**	**97.5**	**77.8**	**65.1**
N8	WinCLIP [32]	89.6	91.4	78.2	96.0	70.2	49.4
	InCTRL [93]	94.3	96.5	91.4	94.9	77.9	59.2
	ResAD [77]	83.6	83.0	75.1	95.5	76.2	47.3
	FoundAD [84]	**94.6**	93.9	90.2	95.1	**78.5**	60.6
	RegAD [29]	92.2	93.0	91.9	**96.5**	76.6	53.3
	PromptAD [36]	91.2	92.6	90.9	94.1	76.1	61.8
A4	DRA [18]	93.0	92.2	91.8	93.9	74.1	61.3
	AHL [91]	93.3	92.0	92.4	94.5	75.8	61.9
N8A4	NAGL [73]	94.0	**97.6**	**93.8**	**96.5**	77.7	**62.8**
	IDEAL (ours)	**96.2**	**98.1**	**93.9**	**97.7**	**78.9**	**66.3**

Table A8: **Comparison results on BraTS** under various few-shot AD settings. Best results are in **red bold**, with second best results in **blue bold**.

SETUPS	METHODS	IMAGE-LEVEL			PIXEL-LEVEL		
		AUROC	AP	F1-Max	AUROC	PRO	F1-Max
N1	WinCLIP [32]	68.8	89.4	90.9	91.0	67.6	25.9
	InCTRL [93]	69.5	89.2	90.3	93.1	70.0	**36.7**
	ResAD [77]	59.7	83.1	90.6	89.5	63.9	23.8
	FoundAD [84]	62.7	**90.3**	90.7	87.8	65.3	29.4
	RegAD [29]	64.2	88.3	**91.6**	92.2	72.8	35.1
	PromptAD [36]	68.7	89.1	90.7	91.9	70.9	34.7
A1	DRA [18]	67.2	83.9	90.6	87.5	71.6	28.7
	AHL [91]	**70.4**	88.3	91.1	90.4	71.1	32.73
N1A1	NAGL [73]	67.7	88.5	91.2	**93.7**	**75.2**	33.8
	IDEAL (ours)	**74.9**	**90.2**	**91.4**	**94.8**	**76.3**	**40.1**
N2	WinCLIP [32]	74.8	**92.4**	91.1	92.7	72.0	30.8
	InCTRL [93]	**75.0**	91.0	90.3	94.1	74.9	**41.3**
	ResAD [77]	65.4	86.7	90.8	92.2	69.6	35.4
	FoundAD [84]	72.5	90.8	91.1	91.9	69.4	33.1
	RegAD [29]	70.7	89.1	**91.8**	94.1	73.4	39.5
	PromptAD [36]	74.4	90.5	91.0	93.7	74.2	41.1
N2A1	NAGL [73]	74.6	91.1	91.7	**94.3**	**77.2**	39.1
	IDEAL (ours)	**78.5**	**93.4**	**91.9**	**95.6**	**77.4**	**43.9**
N4	WinCLIP [32]	75.7	92.4	91.5	93.2	72.6	32.8
	InCTRL [93]	**77.8**	**93.2**	91.2	93.8	75.8	41.6
	ResAD [77]	69.6	87.8	91.7	93.0	72.2	40.3
	FoundAD [84]	73.5	91.2	91.7	92.8	72.6	39.3
	RegAD [29]	75.8	89.6	**92.0**	**94.5**	76.5	40.1
	PromptAD [36]	74.5	91.1	90.7	94.0	74.9	**42.3**
N4A1	NAGL [73]	74.9	91.5	91.8	94.4	**78.6**	41.8
	IDEAL (ours)	**80.5**	**94.1**	**92.1**	**95.8**	**79.2**	**46.6**
N8	WinCLIP [32]	76.9	92.6	92.0	93.8	73.1	33.1
	InCTRL [93]	**79.5**	**93.9**	91.6	94.9	76.6	**47.1**
	ResAD [77]	71.7	89.2	91.9	94.3	73.2	42.4
	FoundAD [84]	76.7	91.8	**92.1**	93.5	73.4	40.8
	RegAD [29]	76.3	89.8	92.0	95.0	**79.2**	45.2
	PromptAD [36]	75.6	92.8	91.0	94.6	76.4	43.9
A4	DRA [18]	73.5	87.6	91.7	89.9	72.7	38.1
	AHL [91]	72.9	91.1	91.3	90.3	75.2	35.4
N8A4	NAGL [73]	75.1	92.4	92.0	**95.2**	78.9	41.3
	IDEAL (ours)	**82.8**	**94.2**	**92.4**	**96.9**	**79.5**	**53.1**

Table A9: **Comparison results on Liver** under various few-shot AD settings. Best results are in **red bold**, with second best results in **blue bold**.

SETUPS	METHODS	IMAGE-LEVEL			PIXEL-LEVEL		
		AUROC	AP	F1-Max	AUROC	PRO	F1-Max
N1	WinCLIP [32]	59.5	57.2	63.2	**96.2**	**92.0**	21.9
	InCTRL [93]	59.7	55.1	60.2	96.0	90.8	22.1
	ResAD [77]	53.3	55.8	57.8	92.7	88.3	11.4
	FoundAD [84]	58.7	57.1	63.7	93.8	89.1	19.2
	RegAD [29]	59.0	**57.9**	61.9	94.0	85.1	17.2
	PromptAD [36]	58.3	54.6	**65.6**	94.3	83.4	18.5
A1	DRA [18]	56.4	54.4	63.7	93.5	83.7	**24.5**
	AHL [91]	**60.2**	55.8	64.1	93.7	80.1	21.6
N1A1	NAGL [73]	59.9	57.6	64.0	94.8	87.9	17.5
	IDEAL (ours)	**62.7**	**57.8**	**65.9**	**97.2**	**93.5**	**25.6**
N2	WinCLIP [32]	61.6	57.6	66.6	**96.5**	**92.1**	22.4
	InCTRL [93]	63.7	58.7	**66.8**	96.1	90.8	**26.5**
	ResAD [77]	58.6	60.3	64.6	92.9	89.5	15.8
	FoundAD [84]	63.9	59.7	65.9	94.0	90.3	21.6
	RegAD [29]	62.1	58.4	62.5	94.2	85.7	17.7
	PromptAD [36]	62.5	58.6	66.1	94.4	84.2	19.7
N2A1	NAGL [73]	**67.2**	**66.3**	64.5	95.0	90.3	18.6
	IDEAL (ours)	**72.4**	**70.3**	**68.1**	**97.7**	**94.9**	**26.9**
N4	WinCLIP [32]	63.7	58.0	67.4	**96.8**	**92.5**	24.1
	InCTRL [93]	67.9	63.2	**69.5**	96.3	92.1	**27.2**
	ResAD [77]	65.4	64.5	67.3	94.7	91.2	18.1
	FoundAD [84]	68.7	64.1	66.0	94.2	90.6	21.9
	RegAD [29]	69.6	59.3	63.2	95.0	89.3	25.2
	PromptAD [36]	68.9	59.1	68.1	94.2	84.9	20.1
N4A1	NAGL [73]	**70.4**	**67.7**	68.0	95.2	91.4	19.1
	IDEAL (ours)	**73.9**	**70.8**	**71.5**	**97.7**	**95.3**	**28.4**
N8	WinCLIP [32]	65.1	58.4	68.6	**96.8**	**93.1**	26.2
	InCTRL [93]	68.1	64.8	70.0	96.5	92.4	**27.4**
	ResAD [77]	69.7	65.1	67.6	96.5	91.8	19.6
	FoundAD [84]	70.9	67.0	69.2	94.3	91.1	23.4
	RegAD [29]	73.7	61.1	63.5	95.3	90.4	27.3
	PromptAD [36]	73.5	70.2	70.7	94.9	84.7	21.7
A4	DRA [18]	70.1	69.9	65.9	93.8	85.9	25.2
	AHL [91]	72.2	68.5	69.7	93.1	89.7	23.9
N8A4	NAGL [73]	**74.9**	**73.9**	**70.8**	95.5	92.5	19.7
	IDEAL (ours)	**77.4**	**74.0**	**71.7**	**97.9**	**95.5**	**30.1**

Table A10: **Comparison results on RESC** under various few-shot AD settings. Best results are in **red bold**, with second best results in **blue bold**.

SETUPS	METHODS	IMAGE-LEVEL			PIXEL-LEVEL		
		AUROC	AP	F1-Max	AUROC	PRO	F1-Max
N1	WinCLIP [32]	69.2	65.9	64.0	90.8	72.8	42.1
	InCTRL [93]	**80.0**	72.8	**74.5**	91.5	**75.8**	44.2
	ResAD [77]	64.1	57.7	61.7	80.1	59.2	28.3
	FoundAD [84]	78.5	73.2	72.9	87.1	61.1	45.9
	RegAD [29]	76.1	72.6	71.2	88.3	69.8	41.9
	PromptAD [36]	77.4	75.3	68.6	91.2	69.3	**46.8**
A1	DRA [18]	76.4	74.7	69.6	86.4	69.8	43.9
	AHL [91]	75.5	74.6	70.4	85.8	68.9	44.9
N1A1	NAGL [73]	78.9	**75.9**	71.6	**91.7**	73.4	46.6
	IDEAL (ours)	**82.8**	**79.2**	**75.9**	**92.5**	**77.0**	**48.1**
N2	WinCLIP [32]	73.7	66.0	64.5	90.9	73.6	42.9
	InCTRL [93]	80.3	74.8	**75.0**	92.1	**76.8**	46.8
	ResAD [77]	69.3	60.5	62.4	81.3	63.1	31.4
	FoundAD [84]	79.0	77.1	73.8	90.2	62.7	48.3
	RegAD [29]	77.4	73.4	73.6	89.7	71.2	46.4
	PromptAD [36]	78.2	75.6	68.7	91.6	70.5	47.2
N2A1	NAGL [73]	**80.4**	**81.7**	73.7	**92.2**	75.6	**49.1**
	IDEAL (ours)	**84.4**	**83.2**	**76.4**	**93.5**	**77.8**	**50.7**
N4	WinCLIP [32]	75.2	70.5	68.4	91.3	74.1	45.8
	InCTRL [93]	**86.5**	79.7	**80.2**	**92.8**	**79.2**	49.1
	ResAD [77]	73.7	67.1	62.7	85.3	65.2	35.3
	FoundAD [84]	81.5	**83.1**	74.4	90.8	68.3	49.2
	RegAD [29]	83.3	75.9	74.2	91.4	74.6	47.3
	PromptAD [36]	81.3	75.7	68.9	92.2	71.5	48.1
N4A1	NAGL [73]	84.8	82.0	75.6	92.6	79.0	**50.7**
	IDEAL (ours)	**89.4**	**85.6**	**81.3**	**94.9**	**79.4**	**51.2**
N8	WinCLIP [32]	76.7	75.6	69.4	92.4	74.6	46.3
	InCTRL [93]	**86.8**	80.7	**81.0**	93.6	**79.6**	50.1
	ResAD [77]	74.8	71.1	63.3	87.6	69.3	38.6
	FoundAD [84]	84.5	83.6	78.3	91.3	69.5	**52.1**
	RegAD [29]	83.9	78.5	75.9	92.6	72.8	49.8
	PromptAD [36]	83.3	76.1	72.8	93.2	73.9	49.3
A4	DRA [18]	82.1	78.4	70.1	89.3	71.2	50.4
	AHL [91]	85.7	78.9	70.4	89.8	73.4	49.9
N8A4	NAGL [73]	85.2	**85.1**	77.6	**94.0**	79.2	51.4
	IDEAL (ours)	**90.6**	**86.8**	**83.7**	**95.6**	**79.9**	**56.8**

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our contributions are summarized in the Abstract and Introduction (Sec. 1) sections. The experiments (Sec. 5) also confirm our contributions.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Please see Appendix G for limitations.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: Please see Appendix B for the generalization analysis and proof.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Please see Implementation Details in Sec. 5 and Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The code is provided in the supplementary material. Our code and data are available at https://github.com/mala-lab/IDEAL.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Please see Sec. 5.1 and Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: As mentioned in Sec. 5.1, we report results averaged across three independent runs with different random seeds.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Please see Implementation Details in Sec. 5.1 and Efficiency Comparison in Sec. 5.3.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics https://neurips.cc/public/EthicsGuidelines?

   Answer: [Yes]

   Justification: We have read NeurIPS Code of Ethics and confirm that our research complies with these ethical standards.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: Please see Appendix G.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: We think our work poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: All datasets used in this paper are all open-source, and we have cited the corresponding papers.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: Our work proposes a novel few-shot anomaly detection framework without releasing new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: Our paper do not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: Our paper do not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The LLM is used only for editing and polishing the manuscript.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.