This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

FastRef: Fast Prototype Refinement for Few-shot Industrial Anomaly Detection

Yufei Li[1] Long Tian[1,†] Yuyang Dai[1] Wenchao Chen[2,†] Liang Bao[1] Xiyang Liu[1]
{1. School of Computer Science and Technology, 2. School of Electronic Engineering}, Xidian University
{tianlong, chenwenchao}@xidian.edu.cn

Abstract

Few-shot industrial anomaly detection (FS-IAD) presents a critical challenge for practical automated inspection systems operating in data-scarce environments. While existing approaches predominantly focus on obtaining prototypes from limited normal images, they neglect to systematically incorporate statistics of query image to enhance prototype representativeness. To address this issue, we propose FastRef, a novel and efficient prototype refinement framework for FS-IAD. Our method operates through an iterative updates during inference: (1) characteristic transfer from query features to the enhanced prototypes, and (2) anomaly suppression by aligning the enhanced prototypes with their normal counterparts. The characteristic transfer is achieved through linear reconstruction of query features from prototypes with an optimizable transport matrix, while the anomaly suppression addresses a key observation in FS-IAD that unlike conventional IAD with abundant normal images, the limited-sample setting makes anomaly reconstruction more probable in characteristic transfer. Therefore, we employ optimal transport to measure while minimize the gap between normal prototypes and their enhanced counterparts for anomaly suppression. For comprehensive evaluation, we integrate FastRef with three competitive prototype-based FS-IAD methods: PatchCore, WinCLIP, and AnomalyDINO. Extensive experiments across four benchmark datasets of MVTec, ViSA, MPDD and RealIAD demonstrate both the effectiveness and efficiency of our approach under 1/2/4-shots. Code will be available at https://github.com/liyufei25/FastRef.

1. Introduction

Industrial Anomaly Detection (IAD) aims to automatically identify defects on product surfaces [13] and has been attracting tremendous attention [15, 32, 34]. The fragmented nature of industrial anomalies ranging from subtle bruises to obvious breakages, with varying appearances and scales [21] , making it difficult for fully supervised methods [7, 11]. Therefore, unsupervised IAD methods, trained with massive normal images, have been developed [14, 16]. In the cold start problem, it is not always possible to obtain a large number of normal images for different products, making existing methods less effective in low data regime [8].

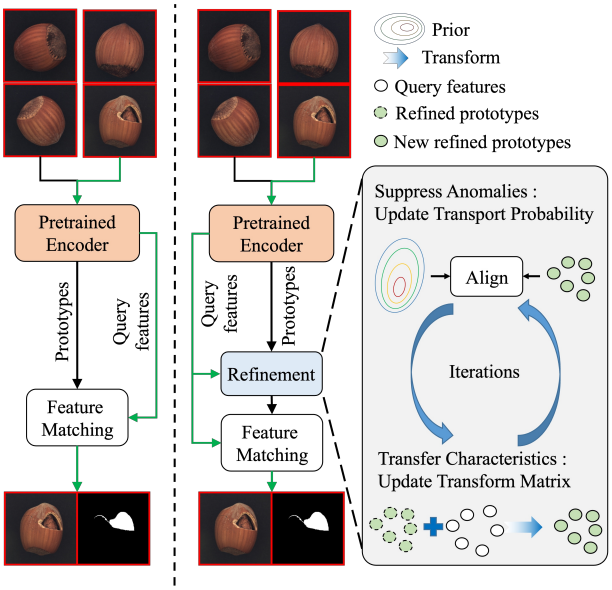


Figure 1. The black line represents the processing flow of supporting images, and the green line represents the processing flow of query images. **Left:** Traditional prototype-oriented FS-IAD methods. **Right:** Our proposed FastRef, it refines prototypes with the query features by suppressing anomalies and transferring characteristics iteratively. The improved performance can be obtained thanks to the well enhanced prototypes in low data regime.

Few-shot learning (FSL) methods [24, 25, 29] have been introduced to unsupervised IAD, allowing the development of a common model generalizing to new products with only a few normal images, such as 1-shot per product. This new paradiagm is known as few-shot IAD [8] (FS-IAD), which mainly involves prototype-oriented methods [5, 9]. At training time, these methods typically use the statistics of few normal support images to construct normal prototypes. At

† Corresponding Author

test time, anomaly scores are computed by measuring the differences between query image and normal prototypes using measurement function. Finally, anomalies are detected and localized by comparing the anomaly scores against the predefined threshold. Especially, [5] further uses query features to quickly refine prototypes at test time. However, we find that point-to-point reconstruction as in [5] does significantly limits the ability to transfer characteristics from query images to prototypes. Additionally, meta learning based FS-IAD methods [8, 30] have been introduced to achieve fast generalization, whose performance is verified to be far behind of prototype-oriented methods [31].

Observing the fact that characteristics of query images have not been fully explored during inference previously [5, 22] either from the data perspective [22] or from the optimization perspective [5], which may result in suboptimal enhanced prototypes , particularly when the available normal images are extremely limited. As shown in Fig. 1, we therefore propose FastRef, a fast prototype refinement model which can be easily integrated with other prototype-oriented FS-IAD methods. FastRef enhances the representativeness of prototypes by transferring characteristics from query features to the enhanced prototypes while suppressing anomalies by aligning the enhanced prototypes with their normal counterparts, making the enhanced prototypes more robust and generalizable. We formulate prototype refinement with query features as a nested optimization process balancing anomaly suppression and characteristic transfer, with a transport probability and transform matrix capturing these two behaviors, simultaneously. We then develope an efficient algorithm to solve the transport probability and transform matrix iteratively. Specifically, as for anomaly suppression, we use Sinkhorn algorithm [3], an entropy-regularized optimal transport, to align the distribution between the original normal prototypes and their enhanced counterparts, ensuring that the enhanced prototypes are unaffected by anomalies. When it comes to characteristic transfer, we maximize the transferability from query features to the enhanced prototypes by minimizing the reconstruction loss between the enhanced prototypes and their normal counterparts with gradient descent in a closed form. Finally, we integrate FastRef into three popular prototype-oriented FS-IAD methods, PatchCore [21], WinCLIP [9], and AnomalyDINO [4] . Our integrated models consistently improves the performance against their corresponding FS-IAD methods by significant margins across four widely used datasets, including MVTec [1], ViSA [36], MPDD [10], and RealIAD [28]. Importantly, we also find that the proper initialization of transform matrix achieves optimal efficiency-performance trade-offs and only requires less than 10 inner loops by Sinkhorn algorithm, guaranteeing the real-time efficiency of our FastRef in practice. The main contributions can be summarized as follows:

- We propose FastRef, a fast prototype refinement model, to transfer characteristics of query features while suppressing anomalies in query features, which can be integrated with other prototype-oriented FS-IAD methods.
- We formulate FastRef as a nested optimization problem and introduce a fast algorithm to solve the problem precisely in a closed form during inference.
- The experimental results confirm that FastRef is effective in improving FS-IAD performance while maintaining real-time efficiency on the datasets of MVTec, ViSA, MPDD, and RealIAD under 1/2/4-shots.

2. Background

2.1. Task Formulation

We formally define the one-class FS-IAD, adhering to the standard FSL. The model is fine-tuned using k normal support images $\boldsymbol{x}^{\mathrm{s}}_{1:\mathrm{k}}$ and predicts whether the t-th query image $\boldsymbol{x}^{\mathrm{q}}_{\mathrm{t}}$ is anomalous at both the pixel and image levels. In this paper, we fully leverage the pre-trained backbone $f_{\boldsymbol{\theta}^*}$ like ResNet-50 [6] or DINO-V2 [17], parameterized by $\boldsymbol{\theta}^*$, to extract features from both support and query images following the previous FS-IAD methods [4, 9, 21], denoted as:

$$\boldsymbol{f}^{\mathrm{s}} = f_{\boldsymbol{\theta}^*}(\boldsymbol{x}^{\mathrm{s}}_{1:\mathrm{k}}), \quad \boldsymbol{f}^{\mathrm{q}}_{\mathrm{t}} = f_{\boldsymbol{\theta}^*}(\boldsymbol{x}^{\mathrm{q}}_{\mathrm{t}}) \tag{1}$$

where $\boldsymbol{f}^{\mathrm{s}} \in \mathbb{R}^{(\mathrm{k}\times\mathrm{h}\times\mathrm{w})\times\mathrm{c}}$ are normal features, and $\boldsymbol{f}^{\mathrm{q}}_{\mathrm{t}} \in \mathbb{R}^{(\mathrm{h}\times\mathrm{w})\times\mathrm{c}}$ is a query image's features. $\boldsymbol{f}^{\mathrm{s}}$ tend to be redundant, we use Coreset [23] to construct prototypes $\boldsymbol{\mathcal{M}}_{\mathrm{s}} \in \mathbb{R}^{(\alpha\times\mathrm{k}\times\mathrm{h}\times\mathrm{w})\times\mathrm{c}}$ by selecting the most representative normal features from $\boldsymbol{f}^{\mathrm{s}}$ with a downsampling ratio $\alpha \in (0, 1)$. For simplicity, we denote $\mathrm{m} = \mathrm{h} \times \mathrm{w}$ and $\mathrm{n} = \alpha \times \mathrm{k} \times \mathrm{h} \times \mathrm{w}$ for the reminder of this paper.

2.2. Optimal Transport

Although optimal transport (OT) has a rich theory, we limit our discussion to OT for discrete distributions and refer the readers to [19] for details. Let us consider P and Q as two discrete probability distributions on the arbitrary space $\boldsymbol{X} \in \mathbb{R}^{\mathrm{c}}$, which can be formulated as $P = \sum_{\mathrm{i}=1}^{\mathrm{m}} a_{\mathrm{i}}\delta_{\boldsymbol{x}_{\mathrm{i}}}$, and $Q = \sum_{\mathrm{j}=1}^{\mathrm{n}} b_{\mathrm{j}}\delta_{\boldsymbol{y}_{\mathrm{j}}}$. In this case, $\boldsymbol{a} \in \Sigma^{\mathrm{m}}$ and $\boldsymbol{b} \in \Sigma^{\mathrm{n}}$, where Σ^{m} denotes the probability simplex of $\mathbb{R}^{\mathrm{m}}$. The OT distance between $\boldsymbol{a}$ and $\boldsymbol{b}$ can be defined as:

$$\mathrm{OT}(P, Q) = \min_{\boldsymbol{T}\in U(P,Q)} \langle \boldsymbol{T}, \boldsymbol{C} \rangle \tag{2}$$

where $\langle\cdot,\cdot\rangle$ denotes the Frobenius dot-product, $\boldsymbol{C} \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ is the cost matrix with element $C_{\mathrm{i,j}} = \boldsymbol{C}(\boldsymbol{x}_{\mathrm{i}}, \boldsymbol{y}_{\mathrm{j}})$, and $\boldsymbol{C}(\cdot,\cdot)$ is usually distance function like Euclidean or cosine. $\boldsymbol{T} \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ is the doubly stochastic transport probability such that $U(P, Q) := \{\boldsymbol{T} | \sum_{\mathrm{i}=1}^{\mathrm{m}} \boldsymbol{T}_{\mathrm{i,j}} = \boldsymbol{b}_{\mathrm{j}}, \sum_{\mathrm{j}=1}^{\mathrm{n}} \boldsymbol{T}_{\mathrm{i,j}} = \boldsymbol{a}_{\mathrm{i}}\}$. To relax the time-consuming problem when minimizing the OT distance, [3] introduced the entropic regularization, $H = -\sum_{\mathrm{i,j}} \boldsymbol{T}_{\mathrm{i,j}} \ln \boldsymbol{T}_{\mathrm{i,j}}$, leading to the widely-used Sinkhorn algorithm for discrete OT problems.

3. Method

We first formulate FastRef as a nested optimization problem between characteristic transfer and anomaly suppression from query features to the enhanced prototypes. Then we introduce a fast and efficient algorithm to solve FastRef by updating transform matrix and transport probability iteratively. Finally, we implement the industrial anomaly detection and localization through reconstruction using query features and the enhanced prototypes.

3.1. Modeling of FastRef

We aim to refine prototypes by transferring characteristics while suppressing anomalies from query images. We propose a nested optimization for modeling such behaviors as:

$$\boldsymbol{W}^*, \boldsymbol{T}^* = \underset{\boldsymbol{W},\boldsymbol{T}}{\arg\min}\, \mathrm{dis}(\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}, \boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}}) + \lambda \mathrm{OT}(P, Q) \quad (3)$$

where $\mathrm{dis}(\cdot,\cdot)$ is the point-wise distance between two sets like Euclidean, with its specific form dependent on the follow-up applications in Sec. 4. $\boldsymbol{W} \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ is the transform matrix used to transfer the characteristics of query features $\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}$ for obtaining the enhanced prototypes $\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w}} = \boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}}$, where $\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w}} \in \mathbb{R}^{\mathrm{m}\times\mathrm{c}}$. Notably, different from prototype refinement that takes place along the feature dimension, we refer to our transform as composition refinement, the meaning of which is reconstructing query features by selecting items from the normal prototypes $\boldsymbol{\mathcal{M}}_{\mathrm{s}}$. $\boldsymbol{T} \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ is the transport probability having been discussed in Sec.2. It is used to suppress anomalies potentially present in query features $\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}$ by minimizing $\mathrm{OT}(P, Q)$. P and Q are distributions expanded by $\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w}}$ and $\boldsymbol{\mathcal{M}}_{\mathrm{s}}$, where $P = \sum_{\mathrm{i}=1}^{\mathrm{m}} \frac{1}{\mathrm{m}} \delta_{\boldsymbol{\mathcal{M}}_{\mathrm{s,i}}^{\mathrm{w}}}$, $Q = \sum_{\mathrm{j}=1}^{\mathrm{n}} \frac{1}{\mathrm{n}} \delta_{\boldsymbol{\mathcal{M}}_{\mathrm{s,j}}}$, and $\boldsymbol{\mathcal{M}}_{\mathrm{s,i}}, \boldsymbol{\mathcal{M}}_{\mathrm{s,j}}^{\mathrm{w}} \in \mathbb{R}^{\mathrm{c}}$. λ is the balanced coefficient.

Remark: Minimizing the second term in Eq. 3 makes the enhanced prototypes $\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w}}$ much more similar with the normal prototypes $\boldsymbol{\mathcal{M}}_{\mathrm{s}}$, thus suppressing potential anomalies in query features. To better understand this, let us consider a case that P contains anomaly features as shown in Fig. 2, where the detailed expressions on $\boldsymbol{W}$ and $\boldsymbol{T}$ will be discussed later, and we omit the index t for simplicity. With the constraint of $\sum_{\mathrm{j}} \boldsymbol{T}_{\mathrm{i,j}} = \frac{1}{\mathrm{m}}$, although the transport probability may pay attention to the anomaly features in $\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w}}$, in order to minimize the second term in Eq. 3, in the next round of optimization on $\boldsymbol{W}$, theoretically, it may ignore those anomaly features by adjusting the corresponding weights in $\boldsymbol{W}$. Therefore, the anomalies could be suppressed. On the contrary, if we only adopt point-wise distance like Euclidean between $\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}$ and $\boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}}$ as shown in the first term of Eq. 3. Then the enhanced prototypes may be misled by anomalies in $\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}$ so that the distance between $\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}$ and $\boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}}$ can be smaller. Additionally, the OT-based

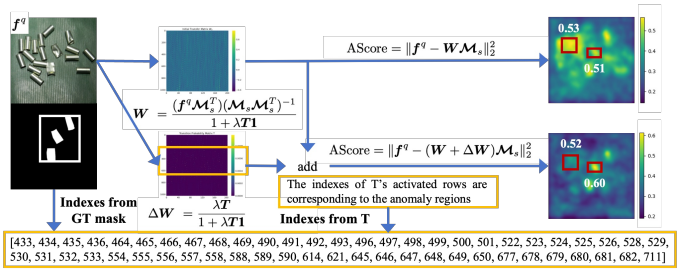


Figure 2. The roles of transform matrix $\boldsymbol{W}$ and transport probability $\boldsymbol{T}$. **Top:** $\boldsymbol{W}$-based $\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}$ reconstruction restores normal features but may propagate anomalies, causing similar anomaly score (AScore) in defective and normal regions. **Bottom:** $\boldsymbol{T}$-adjusted $\boldsymbol{W}$ enables effective anomaly suppression, making defective regions' AScore distinguishable. The indexes between anomaly GT mask and $\boldsymbol{T}$ show that $\boldsymbol{T}$ primarily activates on anomalous areas.

regularization in Eq. 3 does not rely on Gaussian assumption of (enhanced) prototypes, making it more adaptable for real-world FS-IAD, as discussed in the Appendix F.

We then use Sinkhorn algorithm [3] and rewrite the second term in Eq. 3 as follows:

$$\mathrm{OT}_{\epsilon}(P_{\mathrm{s}}, Q_{\mathrm{s}}) := \sum_{\mathrm{i,j}}^{\mathrm{m,n}} \boldsymbol{C}_{\mathrm{i,j}} \boldsymbol{T}_{\mathrm{i,j}} + \epsilon \sum_{\mathrm{i,j}}^{\mathrm{m,n}} \boldsymbol{T}_{\mathrm{i,j}} \ln \boldsymbol{T}_{\mathrm{i,j}} \quad (4)$$

where $\epsilon > 0$ is another balanced coefficient, $\boldsymbol{C} \in \mathbb{R}^{m\times n}$ is typically formulated with simple distance function $\mathrm{dis}(\cdot,\cdot)$ like Euclidean. Recalling that $\boldsymbol{T} \in \mathbb{R}^{\mathrm{m}\times\mathrm{n}}$ must satisfy $U(P, Q) := \{\sum_{\mathrm{j}=1}^{\mathrm{n}} \boldsymbol{T}_{\mathrm{i,j}} = \frac{1}{\mathrm{m}}, \sum_{\mathrm{i}=1}^{\mathrm{m}} \boldsymbol{T}_{\mathrm{i,j}} = \frac{1}{\mathrm{n}}\}$ in FS-IAD, where $\boldsymbol{T}_{\mathrm{i,j}}$ denotes the transport probability between the i-th enhanced prototype and the j-th prototype, serving as an upper-bounded positive metric. Consequently, $\boldsymbol{T}$ naturally weights the importance of each enhanced prototype in relation the set of original normal prototypes.

3.2. Solving of FastRef

It is obvious that the optimal parameters for the transport probability $\boldsymbol{T}^*$ and the transform matrix $\boldsymbol{W}^*$ are interdependent accoridng to Eq. 3, which motivates us to develop an iterative algorithm for solving FastRef. Conceptually, at the l-th iteration, for the purpose of anomaly suppression, we keep the transform matrix $\boldsymbol{W}_l$ fixed and update the transport probability $\boldsymbol{T}_l$ by minimizing the second term of Eq. 3 using Sinkhorn algorithm [3] to derive $\boldsymbol{T}_{l+1}$. For characteristic transfer, we keep the transport probability $\boldsymbol{T}_{l+1}$ fixed and update transform matrix $\boldsymbol{W}_l$ by minimizing Eq. 3, denoted as $\mathcal{L}(\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}, \boldsymbol{\mathcal{M}}_{\mathrm{s}}; \boldsymbol{W}, \boldsymbol{T})$, using gradient descent denoted by $\boldsymbol{W}_{l+1} = \boldsymbol{W}_l + \beta \frac{\partial \mathcal{L}(\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}, \boldsymbol{\mathcal{M}}_{\mathrm{s}}; \boldsymbol{W}_l, \boldsymbol{T}_{l+1})}{\partial \boldsymbol{W}_l}$, β is learning rate. After L steps, the enhanced prototypes with optimal representativeness can be expressed as $\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w}*} = \boldsymbol{W}^* \boldsymbol{\mathcal{M}}_{\mathrm{s}}$, where $\boldsymbol{W}^* = \boldsymbol{W}_{\mathrm{L}}$ and $\boldsymbol{T}^* = \boldsymbol{T}_{\mathrm{L}}$. When we choose $\mathrm{dis}(\cdot,\cdot)$ in Eq. 3 as Euclidean or cosine

distance, $\boldsymbol{W}_{l+1}$ has a closed-form solution, and we have:

$$\boldsymbol{W}_{l+1} = \frac{(\boldsymbol{f}_t^q \boldsymbol{\mathcal{M}}_s^T + \lambda \boldsymbol{T}_l \boldsymbol{\mathcal{M}}_s \boldsymbol{\mathcal{M}}_s^T)(\boldsymbol{\mathcal{M}}_s \boldsymbol{\mathcal{M}}_s^T)^{-1}}{1 + \lambda \boldsymbol{T}_l \cdot \mathbf{1}} \quad (5)$$

where $\boldsymbol{\mathcal{M}}_{\mathrm{s}} \boldsymbol{\mathcal{M}}_{\mathrm{s}}^T \in \mathbb{R}^{\mathrm{n} \times \mathrm{n}}$, we can compute the inverse of the matrix in advance and prepare to reuse it for improving efficiency. Additionally, the direct computation of pseudo-inverse proved numerically stable in experments. In order to take a balance between accuracy and inference speed, we use Coreset [23] to compress $\boldsymbol{\mathcal{M}}_{\mathrm{s}}$ and set n small. Additionally, updating $\boldsymbol{T}_l$ at each iteration is another important computation bottleneck, we find that proper initialization with $\boldsymbol{W}_0 = (\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}} \boldsymbol{\mathcal{M}}_{\mathrm{s}}^T)(\boldsymbol{\mathcal{M}}_{\mathrm{s}} \boldsymbol{\mathcal{M}}_{\mathrm{s}}^T)^{-1}$ achieves optimal efficiency-performance trade-offs and only requires less than $\mathrm{L}_{\mathrm{OT}} = 10$ inner loops by Sinkhorn algorithm, the real-time efficiency for which is also reported in the Appendix E. We also find that $\mathrm{L} = 2$ yields promising performance, as reported in Table 4. All of these evidences indicate that the developed algorithm for solving FastRef satisfies real-time requirements at test time. The derivation of Eq. 5 is presented in the Appendix B.

3.3. Anomaly Detection and Localization

Once the optimal enhanced prototypes $\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w}*}$ are obtained by solving FastRef, the anomaly score map $\boldsymbol{s}$ for the t-th query image $\boldsymbol{x}_{\mathrm{t}}^{\mathrm{q}}$ can be defined by calculating the similarities between $\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w}*}$ and $\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}$, the features of $\boldsymbol{x}_{\mathrm{t}}^{\mathrm{q}}$, as follows:

$$\boldsymbol{s}_{\mathrm{j}} = \mathrm{dis}(\boldsymbol{f}_{\mathrm{t,j}}^{\mathrm{q}}, \boldsymbol{\mathcal{M}}_{\mathrm{s,j}}^{\mathrm{w}*}), \quad \mathrm{j} = 1, ..., \mathrm{m} \quad (6)$$

where the distance function $\mathrm{dis}(\cdot, \cdot)$ is consistent with that in Eq. 3. For image-level anomaly detection, we represent the maximum score s^* among all values in $\boldsymbol{s} \in \mathbb{R}^{\mathrm{m}}$ as $s^* = \max_{\mathrm{j} \in [1,\mathrm{m}]} \boldsymbol{s}_{\mathrm{j}}$. For pixel-level localization, we first upscale the anomaly score map $\boldsymbol{s}$ using bi-linear interpolation to match the original input resolution. We smooth the score map using a Gaussian kernel following [21]. Additionally, the pseudo-code is provided in Alg. 1.

Remark: Notably, our reconstruction-based anomaly detection demonstrates unique advantages over memory-based and other reconstruction-based approaches, an aspect often overlooked in prior work: i) Structure-aware detection (Eq. 6): Unlike patch-wise matching in memory-based methods ($\boldsymbol{s}_{\mathrm{j}} := \min_{\boldsymbol{r} \in \boldsymbol{\mathcal{M}}_{\mathrm{s}}} \mathrm{dis}(\boldsymbol{f}_{\mathrm{t,j}}^{\mathrm{q}}, \boldsymbol{r})$ for $\mathrm{j} = 1, ..., \mathrm{m}$), our approach leverages the structural relationships between query features and their reconstructions, proving particularly effective in low-data regimes; ii) Parameter efficiency versus overfitting: While methods like UniAD [32] and HVQTrans [15] employ complex decoders to mitigate shortcut learning, they risk overfitting due to excessive parameters. In contrast, our lightweight transform matrix (a minimalistic "decoder") efficiently captures feature differences with minimal parameters (Eq. 6); iii) Built-in shortcut resistance: our OT-based anomaly suppression aligns with the design principles of reconstruction-based methods, further preventing shortcuts without added complexity.

Algorithm 1 FS-IAD of FastRef

Input: Maximum iteration number L, prototypes $\boldsymbol{\mathcal{M}}_{\mathrm{s}}$, the t-th query features $\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}$
Parameter: Initial transform matrix $\boldsymbol{W}_0$
Output: Results of FS-IAD

```
Calculate Q using M_s
while l ≤ L − 1 do
    Calculate P using W_l M_s
    Update T_{l+1} by minimizing OT_ε(P, Q) while fix W_l
    Update W_{l+1} according to Eq. 5 while fix T_{l+1}
end while
Set W* = W_L, M_s^{w*} = W* M_s
Implement FS-IAD according to Eq. 6
return FS-IAD results
```

3.4. Convergence Analysis

The algorithm to solve FastRef by iteratively updating transform matrix and transport probability discussed above is convergent and the following inequality holds:

$$\mathcal{L}(\boldsymbol{W}_{l+1}, \boldsymbol{T}_{l+1}) \leq \mathcal{L}(\boldsymbol{W}_l, \boldsymbol{T}_l) - \frac{\mu}{2} \|\boldsymbol{W}_{l+1} - \boldsymbol{W}_l\|_{\mathrm{F}}^2 \quad (7)$$

where $\mathcal{L}(\cdot, \cdot)$ is the objective same with Eq. 3. Since the objective associated with the transform matrix $\boldsymbol{W}$ is convex (see the Appendix), the convergent speed of FastRef by minimizing $\mathcal{L}(\cdot, \cdot)$ should be quick, which is also verified in our follow-up experiments. The proof of the inequality can be found in the Appendix C.

4. Applications to Prototype-oriented Methods

Our proposed FastRef is applied to three popular prototype-oriented FS-IAD methods, including PatchCore [21], WinCLIP [9], and AnomalyDINO [4].

4.1. PatchCore+: Application to PatchCore

PatchCore [21] is originally introduced for one-class IAD with a large number of normal images. Recently, however, it has also been applied to FS-IAD, thanks to its flexible prototype-oriented design [22]. A notable drawback of PatchCore is that its prototypes remain fixed during inference, causing it to overlook the characteristics of the query image features. Therefore, we propose PatchCore+ by incorporating FastRef to enhance the representativeness of prototypes in PatchCore. Specifically, we use Euclidean distance in line with PatchCore, and substitute it into Eq.

3 and Eq. 6, allowing us to rewrite the nested optimization problem and anomaly score map as:

$$\begin{aligned} &\boldsymbol{W}^*, \boldsymbol{T}^* := \arg\min_{\boldsymbol{W},\boldsymbol{T}} \|\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}} - \boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}}\|^2 + \lambda \mathrm{OT}(P,Q) \\ &\boldsymbol{s}_{\mathrm{j}} := \min_{\boldsymbol{r}\in\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w*}}} \|\boldsymbol{f}_{\mathrm{t,j}}^{\mathrm{q}} - \boldsymbol{r}\|^2, \quad \mathrm{j}=1,...,\mathrm{m} \end{aligned} \tag{8}$$

Anomaly detection and localization can finally be implemented by using reconstruction-based metric in Sec. 3.3.

4.2. WinCLIP+: Application to WinCLIP

Unlike PatchCore whose features are extracted by CNNs [26, 33], WinCLIP [9] extracts features using a CLIP-based visual language model [20]. Although it had been demonstrated superior in FS-IAD, WinCLIP also neglects the importance of efficiently transferring characteristics from query images to enhance the prototypes. We propose WinCLIP+ by applying FastRef to WinCLIP, it uses cosine distance to rewrite Eq. 3 and Eq. 6 as:

$$\begin{aligned} &\boldsymbol{W}^*, \boldsymbol{T}^* := \arg\min_{\boldsymbol{W},\boldsymbol{T}} \frac{1}{2}[1-\cos(\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}, \boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}})] + \lambda \mathrm{OT}_{\epsilon}(P,Q) \\ &\boldsymbol{s}_{\mathrm{j}} := \min_{\boldsymbol{r}\in\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w*}}} \frac{1}{2}[1-\cos(\boldsymbol{f}_{\mathrm{t,j}}^{\mathrm{q}}, \boldsymbol{r})], \quad \mathrm{j}=1,...,\mathrm{m} \end{aligned} \tag{9}$$

where $\boldsymbol{\mathcal{M}}_{\mathrm{s}}^{\mathrm{w*}} = \boldsymbol{W}^*\boldsymbol{\mathcal{M}}_{\mathrm{s}}$. To conduct FS-IAD for $\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}$, we combine the maximum value of $\boldsymbol{s}$ above with the WinCLIP zero-shot anomaly score $s_0 : \mathbb{R}^{\mathrm{c}} \to [0,1]$ [9], which provide complementary information, one from query features and the other from CLIP knowledge by:

$$s^* = \frac{1}{2}[s_0(\boldsymbol{f}_{\mathrm{t}}^{\mathrm{q}}) + \max_{j\in[1,\mathrm{m}]} \boldsymbol{s}_j] \tag{10}$$

Finally, the anomaly detection and localization can be achieved by using reconstruction-based metric in Sec. 3.3.

4.3. AnomalyDINO+: Application to AnomalyDINO

AnomalyDINO [4] employs more scalable foundation Transformer, DINOv2 [17], compared with CLIP-based WinCLIP [9] for FS-IAD above. We propose AnomalyDINO+ by integrating our FastRef with AnomalyDINO. Specifically, we first use cosine distance in Eq. 3 and Eq. 6, and then substitute the normal prototypes (also called the memory bank in AnomalyDINO) and query features collected from AnomalyDINO into Eq. 3 and Eq.6 to obtain the enhanced prototypes at test time per query image. According to the above two steps, the formulations of objective function and anomaly score map are same with those of WinCLIP+. Similar to AnomalyDINO [4], we compute image-level anomaly score s^* by averaging the top 1% values of the anomaly map $\boldsymbol{s}$, namely, $s^* = \mathrm{mean}(\boldsymbol{s}_{0.01}^{\max})$, where $\boldsymbol{s}_{0.01}^{\max}$ is the subset containing the top 1% highest values in $\boldsymbol{s}$. This statistic can be understood as an empirical estimate of the extreme-tail anomaly intensity . Selecting top 1% values effectively suppresses noise while retaining the most informative anomaly evidence.

Table 1. Image AUROC (%) ↑ versus different variants of FastRef (Variant and FastRecon) with different regularization, results are reported on MPDD dataset under 2-shots.

Name	Regularizations	AUROC
FastRef	$\sum_{\mathrm{i,j}} \boldsymbol{T}_{\mathrm{i,j}} \|\boldsymbol{W}_{\mathrm{i},:}\boldsymbol{\mathcal{M}}_{\mathrm{s}} - \boldsymbol{\mathcal{M}}_{\mathrm{s}}(\mathrm{j},:)\|_2^2$	78.2
Variant	$\frac{1}{\mathrm{n}}\sum_{\mathrm{i}=1}^{\mathrm{n}} \boldsymbol{W}_{\mathrm{i},:}\boldsymbol{\mathcal{M}}_{\mathrm{s}} - \frac{1}{\mathrm{m}}\sum_{\mathrm{j}=1=1}^{\mathrm{m}} \boldsymbol{\mathcal{M}}_{\mathrm{s}}(\mathrm{j},:)$	76.9
FastRecon	$\|\boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}} - \boldsymbol{\mu}\|_2^2$	76.4

4.4. Comparison to FastRecon

Similar with FastRef, FastRecon [5] also enhances prototypes by transferring query image features. However, these two methods still differs from each other significantly:

- FastRecon constructs prototypes by $\boldsymbol{\mu} = \frac{1}{k}\sum_{\mathrm{i}=1}^{k} \boldsymbol{f}_{\mathrm{i}}^{\mathrm{s}}$. It is obvious that prototypes are ordered since the shapes of $\boldsymbol{\mu}$ and $\boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}}$ are the same. FastRef employs the Coreset of support features as normal prototypes, which is disordered. We have verified the importance of disorder in FS-IAD and report results in Table 1.
- FastRecon assumes the enhanced prototypes and their normal counterparts are isotropic Gaussian distributions. Thus, it uses Euclidean distance for anomaly suppression by $\|\boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}} - \boldsymbol{\mu}\|_2^2$. FastRef only requires that the features be sampleable which should be more precise and flexible than FasRecon, as shown in Table 1 and Appendix F.

To sum up, FastRecon can be regarded as a special case of our method when we assume prototypes are Gaussian distributed and normal prototypes are ordered. That is, setting $\boldsymbol{T} = \mathrm{I}$ and substituting $\boldsymbol{\mu}$ into our anomaly suppression term $\sum_{\mathrm{i,j}} \boldsymbol{T}_{\mathrm{i,j}} \|\boldsymbol{W}_{\mathrm{i},:}\boldsymbol{\mathcal{M}}_{\mathrm{s}} - \boldsymbol{\mathcal{M}}_{\mathrm{s}}(\mathrm{j},:)\|_2^2$, it degrades into $\|\boldsymbol{W}\boldsymbol{\mathcal{M}}_{\mathrm{s}} - \boldsymbol{\mu}\|_2^2$, which is exactly the form of FastRecon.

5. Experiments

We first evaluate both image and pixel-level performance corresponding to the tasks of anomaly detection and localization to demonstrate the effectiveness of FastRef. Then, ablation studies and hyper-parameters' impacts are implemented to validate the improvements of different components in FastRef. We report the inference time of FastRef in Appendix D verifying its efficiency.

5.1. Experiment Setup

(1) Datasets. We use MVTec [1], VisA [36], MPDD [10], and RealIAD [28] datasets under 1/2/4-shots settings. More details of datasets can be found in the Appendix A.

(2) Competing Methods. We compare our developed PatchCore+, WinCLIP+, and AnomalyDINO+ with most

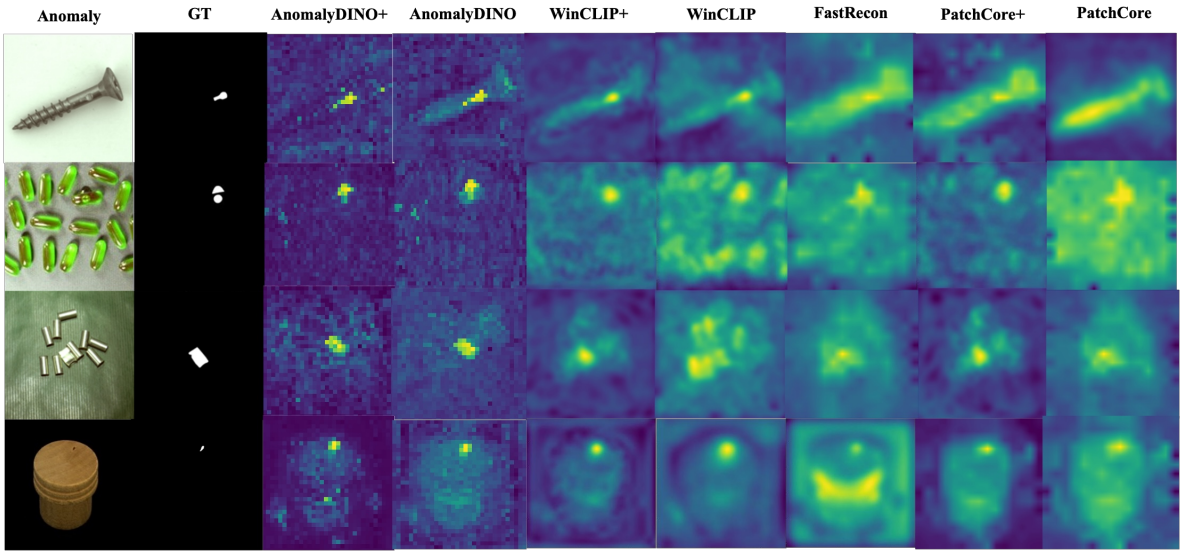


Figure 3. From top to bottom are visualizations of anomaly localization on MVTec, ViSA, MPDD, and RealIAD under 4-shots.

recently proposed FS-IAD methods or those applicable in low-data regimes, including RegAD [8], PatchCore [21], FastRecon [5], WinCLIP [9], AnomalyDINO [4], GraphCore [31], FOCT [27], and PromptAD [12]. For fairness, we report results on all the four datasets by implementing experiments with the official codes of PaDiM, PromptAD, PatchCore, FastRecon, WinCLIP, and AnomalyDINO using the same few normal images. We report results of RegAD, FOCT, and GraphCore from their original papers.

(3) Evaluation Protocols. We evaluate the performance of few-shot anomaly detection and localization separately using Image and Pixel-level AUROC following [9].

(4) Implementation Details. We separately follow the basic settings of PatchCore [21], WinCLIP [9], and AnomalyDINO [4] to implement experiments. For PatchCore, FasRecon and PatchCore+, we use the pretrained WRN-50 [33] to extract features from the intermediate layers, following [21]. Images are all resized to 256×256. Balanced coefficient $\lambda = 0.3$ and Coreset sampling ratio $\alpha = 0.05$ for all the datasets. As for WinCLIP+, we resize all images to 240×240 and use the pre-trained CLIP model with ViT-B/16+ to extract features following [9]. For developing AnomalyDINO+, we use the smallest model of DINOv2 [17], and evaluate input resolution with 448 pixels following [4]. For the last two variants WinCLIP+ and AnomalyDINO+, $\lambda = 0.1$ and $\alpha = 0.5, 0.3, 0.2, 0.1$ for MVTec-AD, VisA, MPDD, and RealIAD datasets. At test time, the batch size equals to 1 to isolate the performance gains from our method. All experiments are conducted on a NVIDIA GTX 3090 GPU. To conveniently verify the effectiveness of the proposed method integrated with AnomalyDINO, we do not employ foreground masking or image rotation.

5.2. Comparisons with SOTA Methods

(1) Image level Comparison. We present anomaly detection results in Table 2. Comparing the results of RegAD [8], PatchCore [22], and FastRecon [5], we find that: i) Prototype-oriented methods outperform the meta-learning based methods, demonstrating superior flexibility of their features; ii) FastRecon outperforms PatchCore, highlighting the importance of incorporating characteristics from query images. When comparing PatchCore and FasRecon to WinCLIP [9], WinCLIP achieves a substantial performance gain except on the MPDD dataset. We attribute this to two factors: i) Generally, the representations from pre-trained CLIP models are more powerful than those from CNNs; ii) Unlike MVTec, VisA and RealIAD, MPDD is a metal dataset with rotation variations that may not be well-represented during CLIP pre-training. We observe consistent performance gain when the scale of pre-trained model increases by comparing the results between WinCLIP and AnomalyDINO. This discrepancy motivates us to demonstrate the effectiveness of our model using pre-trained models with various scale. And there are two key observations: i) Prototype-based FS-IAD methods with pre-trained models of various scale may benefit from our proposed FastRef consistently with minimum modifications; ii) The performance improvement of FastRef is more pronounced in scenarios where the pre-trained model is weaker, the dataset is more challenging, or both factors interact synergistically.

(2) Pixel level Comparison. Pixel-level anomaly localization results are also presented in Table 2. When comparing PatchCore+, WinCLIP+, and AnomalyDINO+ with other methods, the trend in pixel-level AUROC is consistent with that of the image-level AUROC. It suggests that

Table 2. FS-IAD performance on MVTec, MPDD, ViSA, and RealIAD datasets. FR is the abbreviation of FastRef. Δ is improvement. Image- and pixel-level results are reported in AUROC (%) ↑. The best and the second best results are separately bold with black and blue.

Setup	Method	FR	MVTec		MPDD		ViSA		RealIAD	
			Image	Pixel	Image	Pixel	Image	Pixel	Image	Pixel
1-shot	RegAD (ECCV'22)	×	82.9	92.5	60.9	92.6	–	–	–	–
	GraphCore (ICLR'23)	×	89.9	95.6	**84.7**	95.2	–	–	–	–
	FOCT (ACMMM'24)	×	87.1	94.4	**78.9**	96.2	–	–	–	–
	FastRecon (ICCV'23)	×	85.7	93.2	74.1	96.3	76.2	96.7	71.2	95.7
	PromptAD (CVPR'24)	×	92.9	95.1	73.1	95.1	**86.5**	96.2	67.8	91.6
	PatchCore (CVPR'22)	×	84.1	92.3	71.0	96.3	71.0	96.1	70.9	91.2
	PatchCore+ (Ours)	✓	85.9	93.7	74.9	96.6	78.3	**97.1**	**75.9**	**96.3**
	Δ	✓	+1.8	+1.4	+3.9	+0.3	+7.3	+1.0	+5.0	+5.1
	WinCLIP (CVPR'23)	×	93.5	93.6	70.5	96.3	83.4	94.7	73.8	94.3
	WinCLIP+ (Ours)	✓	93.8	95.7	72.5	**96.9**	83.9	95.8	74.4	94.8
	Δ	✓	+0.3	+2.1	+2.0	+0.6	+0.5	+1.1	+0.6	+0.5
	AnomalyDINO (WACV'25)	×	**96.4**	**95.8**	72.0	95.9	82.2	96.3	73.3	95.4
	AnomalyDINO+ (Ours)	✓	**97.0**	**96.5**	76.7	**97.1**	**84.6**	**97.2**	**77.8**	**96.2**
	Δ	✓	+0.6	+0.7	+4.7	+1.2	+2.4	+0.9	+4.5	+0.8
2-shot	RegAD (ECCV'22)	×	85.7	94.6	63.4	93.2	–	–	–	–
	GraphCore (ICLR'23)	×	91.9	**96.9**	**85.4**	95.4	–	–	–	–
	FOCT (ACMMM'24)	×	90.5	94.8	**82.4**	96.5	–	–	–	–
	FastRecon (ICCV'23)	×	88.3	94.5	76.4	96.7	86.1	**97.6**	72.5	95.9
	PromptAD (CVPR'24)	×	93.4	95.4	80.1	95.8	**86.7**	96.5	74.5	93.5
	PatchCore (CVPR'22)	×	87.1	93.3	71.4	96.5	80.0	96.9	71.7	91.3
	PatchCore+ (Ours)	✓	88.8	94.7	78.2	96.9	**87.1**	**98.0**	76.9	**96.5**
	Δ	✓	+1.7	+1.4	+6.8	+0.4	+7.1	+1.1	+5.2	+5.2
	WinCLIP (CVPR'23)	×	93.7	93.8	72.5	96.5	83.8	95.1	75.0	94.6
	WinCLIP+ (Ours)	✓	93.9	96.2	76.0	**97.3**	84.1	96.4	75.9	95.2
	Δ	✓	+0.2	+2.4	+3.5	+0.8	+0.3	+1.3	+0.9	+0.6
	AnomalyDINO (WACV'25)	×	**96.7**	96.0	75.2	96.3	82.5	96.7	**77.8**	96.1
	AnomalyDINO+ (Ours)	✓	**97.2**	**96.7**	78.4	**97.4**	84.8	97.3	**79.8**	**96.8**
	Δ	✓	+0.5	+0.7	+3.2	+1.1	+2.3	+0.6	+2.0	+0.7
4-shot	RegAD (ECCV'22)	×	88.2	95.8	68.3	93.9	–	–	–	–
	GraphCore (ICLR'23)	×	92.9	**97.4**	**85.7**	95.7	–	–	–	–
	FOCT (ACMMM'24)	×	93.2	96.2	**83.2**	96.7	–	–	–	–
	FastRecon (ICCV'23)	×	91.3	96.1	79.7	96.9	88.2	**98.0**	73.2	96.0
	PromptAD (CVPR'24)	×	95.5	96.3	80.4	96.2	**88.8**	96.8	78.5	94.9
	PatchCore (CVPR'22)	×	90.0	95.1	76.2	97.2	84.2	97.5	75.9	93.0
	PatchCore+ (Ours)	✓	92.1	96.1	80.3	97.2	**90.4**	**98.2**	77.4	**96.7**
	Δ	✓	+2.1	+1.0	+4.1	+0.0	+6.2	+0.7	+1.5	+3.7
	WinCLIP (CVPR'23)	×	95.3	94.2	75.0	96.8	84.1	95.4	76.4	94.8
	WinCLIP+ (Ours)	✓	95.5	96.7	82.0	**97.6**	85.0	96.6	77.3	95.3
	Δ	✓	+0.2	+2.5	+7.0	+0.8	+0.9	+1.2	+0.9	+0.5
	AnomalyDINO (WACV'25)	×	**97.1**	96.4	80.3	96.9	86.8	97.1	**79.2**	96.4
	AnomalyDINO+ (Ours)	✓	**97.4**	**96.8**	82.1	**97.6**	87.1	97.9	**81.3**	**96.7**
	Δ	✓	+0.3	+0.4	+1.8	+0.7	+0.3	+0.8	+2.1	+0.3

our model not only effectively detects anomalous images but also accurately localizes the anomalous regions. We attribute these gains to the representativeness of enhanced prototypes with the help of our proposed FastRef. Visualization results of our FastRef-enhanced methods including PatchCore+, WinCLIP+, and AnomalyDINO+ versus

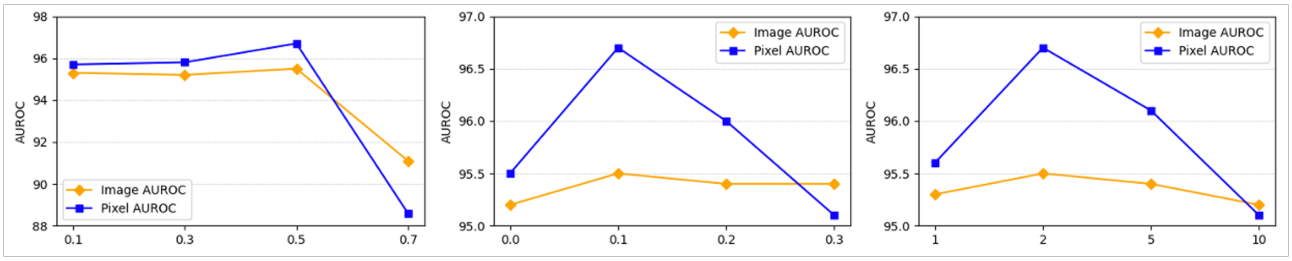


Figure 4. From left to right are results on hyper-parameters of α, λ, and L versus AUROC (%) ↑ using MVTec under 4-shots.

Table 3. Ablation studies of WinCLIP+ with image/pixel-AUROC (%) ↑ under 2-shots. The best results are in bold.

W^*	T^*	MVTec		VisA		MPDD	
		Image	Pixel	Image	Pixel	Image	Pixel
×	×	93.7	93.8	83.8	95.1	72.5	96.5
✓	×	93.8	94.7	84.0	96.2	74.9	96.8
✓	✓	**93.9**	**96.2**	**84.1**	**96.4**	**76.0**	**97.3**

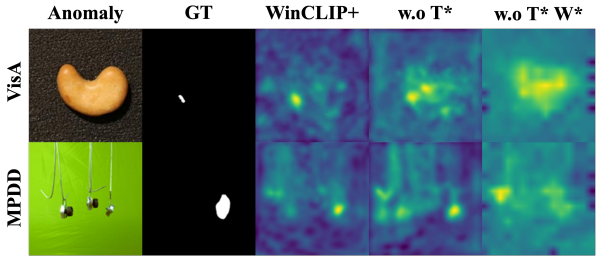


Figure 5. Anomaly localization under 2-shots.

other competitive methods including PatchCore, FasRecon, WinCLIP, and AnomalyDINO in Fig. 3. More visualizations of our methods among different datasets are shown in the Appendix G. Additionally, to avoid any ambiguity or vagueness, it is necessary to provide the clarification that FastRecon+ is the same with PatchCore+ (according to Sec. 4.1), thus we omit reporting results of FastRecon+ in Fig. 3, which is also consistent with Table 2.

5.3. Ablative Analysis

We use WinCLIP+ to assess the impact of transport probability $\boldsymbol{T}^*$ for anomaly suppression and transform matrix $\boldsymbol{W}^*$ for characteristic transfer under 2-shots on the three datasets. The results are reported in Table 3, with visualizations on the datasets of VisA and MPDD are shown in Fig. 5, where w.o $\boldsymbol{T}^*$ means WinCLIP+ without $\boldsymbol{T}^*$, w.o $\boldsymbol{T}^*$&$\boldsymbol{W}^*$ denotes WinCLIP.

(1) Impact of $\boldsymbol{T}^*$. Anomalies in query image negatively affect performance due to the limited diversity and representativeness of normal prototypes. Therefore, leveraging query image through effective anomaly suppression is crucial. As shown in Table 3, anomaly suppression yields averaged improvements of over 0.4% in Image AUROC while over 0.7% in Pixel AUROC on all the three datasets.

(2) Impact of $\boldsymbol{W}^*$. According to Table 3, $\boldsymbol{W}^*$ consistently improves performance. However, in same cases we observe that the gains from using $\boldsymbol{T}^*$ are more pronounced than those from $\boldsymbol{W}^*$. For example, image-level AUROC on MPDD, the gain from using $\boldsymbol{W}^*$ is 2.4%, whereas further using $\boldsymbol{T}^*$ results in a 1.1% improvement. This suggests that relying solely on $\boldsymbol{W}^*$ may introduce anomalies, thereby limiting overall performance improvements.

5.4. Hyper-parameters Analysis

We report results of hyperparameters' impacts using WinCLIP+ under 4-shots setting on MVTec in Fig. 4 , including the Coreset sampling ratio α, the balanced coefficient λ, and the updating iteration L.

(1) Impact of α. We find that α is crucial for determining FS-IAD performance in Fig. 4 (left), since it controls the representativeness of normal prototypes.

(2) Impact of λ. According to the results of Fig. 4 (middle), We observe a similar trend as in α that the performance first improves and then declines. This phenomenon suggests that characteristic transfer should dominate the success of prototype refinement, aligning with our design.

(3) Impact of L. The enhanced prototypes are obtained iteratively, thus, we investigate the impact of L. Results in Fig. 4 (right) indicate that the WinCLIP+ achieves the best performance when $\mathrm{L} = 2$, demonstrating the efficiency of our closed form solution for FastRef.

6. Conclusion

In this work, we propose FastRef, a lightweight prototype refinement module for few-shot industrial anomaly detection. It formulates refinement as a nested optimization with a closed-form solution, enabling efficient characteristic transfer and anomaly suppression. Experiments on four benchmarks show consistent improvements in detection and localization across backbones. FastRef offers a simple, general mechanism for strengthening prototype representations under few-shot constraints and can serve as a useful building block for scalable anomaly detection.

7. Acknowledgement

This work was supported by the National Natural Science Foundation of China (62506277, 62571396, and 82172860), the Fundamental Research Funds for the Central Universities, the Postgraduate Innovation Fund of Xidian University, Shaanxi Province Key R&D Plan(2024GX-ZDCYL-01-11), and Guangxi Unveiling System Science and Technology Project(JB2504240003).

References

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019. 2, 5

[2] Sébastien Bubeck et al. Convex optimization: Algorithms and complexity. *Foundations and Trends® in Machine Learning*, 8(3-4):231–357, 2015.

[3] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013. 2, 3

[4] Simon Damm, Mike Laszkiewicz, Johannes Lederer, and Asja Fischer. Anomalydino: Boosting patch-based few-shot anomaly detection with dinov2. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1319–1329. IEEE, 2025. 2, 4, 5, 6

[5] Zheng Fang, Xiaoyang Wang, Haocheng Li, Jiejie Liu, Qiugui Hu, and Jimin Xiao. Fastrecon: Few-shot industrial anomaly detection via fast feature reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17481–17490, 2023. 1, 2, 5, 6

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2

[7] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017. 1

[8] Chaoqin Huang, Haoyan Guan, Aofan Jiang, Ya Zhang, Michael Spratling, and Yan-Feng Wang. Registration based few-shot anomaly detection. In *European Conference on Computer Vision*, pages 303–319. Springer, 2022. 1, 2, 6

[9] Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. Winclip: Zero-/few-shot anomaly classification and segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19606–19616, 2023. 1, 2, 4, 5, 6

[10] 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 *2021 13th International congress on ultra modern telecommunications and control systems and workshops (ICUMT)*, pages 66–71. IEEE, 2021. 2, 5

[11] Pooja Kamat and Rekha Sugandhi. Anomaly detection for predictive maintenance in industry 4.0-a survey. In *E3S web of conferences*, page 02007. EDP Sciences, 2020. 1

[12] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16838–16848, 2024. 6

[13] Jiaqi Liu, Guoyang Xie, Jinbao Wang, Shangnian Li, Chengjie Wang, Feng Zheng, and Yaochu Jin. Deep industrial image anomaly detection: A survey. *Machine Intelligence Research*, 21(1):104–135, 2024. 1

[14] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20402–20411, 2023. 1

[15] Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection. *arXiv preprint arXiv:2310.14228*, 2023. 1, 4

[16] Ruiying Lu, YuJie Wu, Long Tian, Dongsheng Wang, Bo Chen, Xiyang Liu, and Ruimin Hu. Hierarchical vector quantized transformer for multi-class unsupervised anomaly detection. *Advances in Neural Information Processing Systems*, 36, 2024. 1

[17] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 2, 5, 6

[18] Kaare Brandt Petersen, Michael Syskind Pedersen, et al. The matrix cookbook. *Technical University of Denmark*, 7(15): 510, 2008.

[19] Gabriel Peyré, Marco Cuturi, et al. Computational optimal transport: With applications to data science. *Foundations and Trends® in Machine Learning*, 11(5-6):355–607, 2019. 2

[20] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. 5

[21] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022. 1, 2, 4, 6

[22] João Santos, Triet Tran, and Oliver Rippel. Optimizing patchcore for few/many-shot anomaly detection. *arXiv preprint arXiv:2307.10792*, 2023. 2, 4, 6

[23] Ozan Sener and Silvio Savarese. Active learning for convolutional neural networks: A core-set approach. *arXiv preprint arXiv:1708.00489*, 2017. 2, 4

[24] Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017. 1

[25] Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1199–1208, 2018. 1

[26] Mingxing Tan. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019. 5

[27] Long Tian, Hongyi Zhao, Ruiying Lu, Rongrong Wang, Yu-Jie Wu, Liming Wang, Xiongpeng He, and Xiyang Liu. Foct: Few-shot industrial anomaly detection with foreground-aware online conditional transport. In *ACM Multimedia 2024*. 6

[28] Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jiangning Zhang, Zhihao Gu, Shuguang Qian, Mingang Chen, and Lizhuang Ma. Real-iad: A real-world multi-view dataset for benchmarking versatile industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 22883–22892, 2024. 2, 5

[29] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34, 2020. 1

[30] Jhih-Ciang Wu, Ding-Jie Chen, Chiou-Shann Fuh, and Tyng-Luh Liu. Learning unsupervised metaformer for anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4369–4378, 2021. 2

[31] Guoyang Xie, Jingbao Wang, Jiaqi Liu, Feng Zheng, and Yaochu Jin. Pushing the limits of fewshot anomaly detection in industry vision: Graphcore. *ICLR*, 2023. 2, 6

[32] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems*, 35:4571–4584, 2022. 1, 4

[33] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016. 5, 6

[34] Ying Zhao. Omnial: A unified cnn framework for unsupervised anomaly localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3924–3933, 2023. 1

[35] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International conference on learning representations*, 2018.

[36] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *European Conference on Computer Vision*, pages 392–408. Springer, 2022. 2, 5