# Anomaly-Related Residual Fields for Cross-domain Anomaly Detection

Kewei Gao[1], Jiayi Xie[2], Zhengda Shen[3], Weijun Qin[4], Lingxiang Jia[1],
Kejia Chen[1], Zunlei Feng[1], Yijun Bei[1*]
[1] Zhejiang University, [2] Zhejiang University of Technology,
[3] Chongqing University, [4] EbTech Co. Ltd, China

## Overview

This supplementary material provides comprehensive theoretical and experimental support for the **Residual-Evolution Field (REF)** and **Cross-domain Field Alignment (CFA)** framework. Sections S.0 and S.1 first establish the rigorous notation, basic objects, and assumptions (e.g., local strong log-concavity, Lipschitz score functions, and flow stability). Sections S.2 through S.6 constitute the core theory, demonstrating via theorems how the residual primitives $(R, M, Q)$ decouple from anomalous distributional shifts and achieve dynamic absorption and non-accumulation on normal samples. Finally, Section S.7 provides the theoretical foundation for CFA, proving contraction properties in the REF feature space and deriving the target domain risk bound. Section S.8 further provides experimental details and additional results.

## S.0 Notation and basic objects

**Indices, domains, and norms.** Time indices $t \in \{1, 2, \dots, T\}$ with $T$ the terminal (data) time. Spatial lattice $\Omega = \{1, \dots, H\} \times \{1, \dots, W\}$; a patch is $P \subset \Omega$ with center $u \in \Omega$. Here $H, W \in \mathbb{N}$ denote image height and width, $|P|$ is the pixel-cardinality of patch $P$, and $C \in \mathbb{N}$ will denote the number of channels (used below when specifying the image tensor shape). For a vector $v$, $\|v\|_2$ is the Euclidean norm and $\langle \cdot, \cdot \rangle$ the inner product. For a matrix $A$, $\|A\|_{\mathrm{op}}$ and $\|A\|_F$ denote operator and Frobenius norms, and $(\cdot)^\top$ is transpose. We write $\nabla$ for (vector) gradients with respect to the argument made explicit by context.

**Images, forward perturbation, and score.** An image $x \in \mathbb{R}^{H \times W \times C}$. The forward noising process is

$$y_t = \alpha_t x + \sigma_t \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, I), \quad \alpha_t^2 + \sigma_t^2 = 1,\ \sigma_{t+1} \le \sigma_t, \tag{1}$$

with schedule $(\alpha_t, \sigma_t)$ [5, 9]. Here $I$ is the identity matrix of dimension matching the ambient vectorization of $y_t$. A score network $S_\theta(\cdot, t) \approx \nabla_y \log p_t(y)$ is given at each $t$ [6, 10, 13], where $p_t$ denotes the density of $y_t$ induced by the forward process.

**Same-time projection and reverse flow.** The Tweedie same-time projection is

$$P_t(y) = y + \sigma_t^2\, S_\theta(y, t). \tag{2}$$

This is the classical Tweedie's formula relating the posterior mean and the score under Gaussian noise [4]. Deterministic reverse (probability–flow) dynamics:

$$\dot{z} = g(t)\, S_\theta(z, t), g(t) > 0, \Phi_{t \to T}(y) :\ z(t) = y \mapsto z(T). \tag{3}$$

This ODE is the probability–flow formulation of score-based diffusion [10]. The map $\Phi_{t \to T}$ denotes the solution operator of this ODE from $t$ to $T$.

**Patch aggregation and time differences.** For any per-pixel vector field $X_t(u)$, define the patch mean

$$\bar{X}_t(P) = |P|^{-1} \sum_{u \in P} X_t(u), \tag{4}$$

and the temporal difference $\Delta_t \bar{X}_t(P) = \bar{X}_{t+1}(P) - \bar{X}_t(P)$. Time weights $w_t \ge 0$ satisfy $\sum_{t=1}^{T} w_t = 1$.

**REF primitives (pointwise).** Define for $(u, t)$:

$$R_t(u) = S_\theta(y_t, t)(u) - S_\theta\big(P_t(y_t), t\big)(u), \tag{5}$$

$$v_t(u) = S_\theta\big(P_t(y_t), t\big)(u), \qquad \Pi_v[w] = \tfrac{\langle w, v \rangle}{\|v\|_2^2}\, v,$$

$$M_t(u) = S_\theta(y_t, t)(u) - \Pi_{v_t(u)}\big[S_\theta(y_t, t)(u)\big], \tag{6}$$

$$Q_t(u) = \Phi_{t \to T}(y_t)(u) - \Phi_{t \to T}\big(P_t(y_t)\big)(u). \tag{7}$$

Here $\Pi_v[\cdot]$ is the orthogonal projection onto the direction $v$.

**Patch-level statistics.** Let $\varepsilon_{\mathrm{rid}} > 0$ be a small ridge constant (distinct from the forward noise $\varepsilon$ in (1)). For

*Corresponding author: beiyj@zju.edu.cn

$X \in \{R, M, Q\}$,

$$E_X(P) = \sum_{t=1}^{T} \left\|\bar{X}_t(P)\right\|_2^2 w_t, \tag{8}$$

$$\mathrm{NS}_X(P) = \frac{\sum_{t=1}^{T-1} \left\|\bar{X}_{t+1}(P) - \bar{X}_t(P)\right\|_2^2 w_t}{\sum_{t=1}^{T} \left\|\bar{X}_t(P)\right\|_2^2 w_t + \varepsilon_{\mathrm{rid}}}, \tag{9}$$

$$\mathrm{DV}(P) = \sum_{t=1}^{T-1} \frac{\left\|\bar{M}_{t+1}(P) - \bar{M}_t(P)\right\|_2^2}{\left\|\bar{M}_t(P)\right\|_2^2 + \varepsilon_{\mathrm{rid}}} w_t. \tag{10}$$

**Normal vs. anomalous models.** Let $p_{t,N}$ and $p_{t,A}$ be the normal and anomalous densities and

$$p_t = (1-\pi)\, p_{t,N} + \pi\, p_{t,A}, \tag{11}$$

$$\gamma_A(y,t) = \frac{\pi\, p_{t,A}(y)}{(1-\pi) p_{t,N}(y) + \pi p_{t,A}(y)}. \tag{12}$$

Here $\pi \in [0,1]$ is the mixture proportion; $s_N = \nabla \log p_{t,N}$, $s_A = \nabla \log p_{t,A}$, and $\Delta s = s_A - s_N$.

**Pushforward, Wasserstein, and risk.** For a measurable map $\Psi$ and distribution $P$, $\Psi \# P$ is the pushforward. $W_1(\cdot,\cdot)$ is the 1-Wasserstein distance. For a hypothesis $h$ and bounded loss $\ell \in [0,1]$, the risk on domain $D$ is

$$\varepsilon_D(h) = \mathbb{E}_{(x,y)\sim P_D}\big[\ell(h(x), y)\big]. \tag{13}$$

We write $\mathbb{E}[\cdot]$ for expectation and $\mathrm{Cov}(\cdot)$ for covariance (used later in whitening); Lipschitz maps are non-expansive for $W_1$ [12].

## S.1 Assumptions

**A1 (Local strong log-concavity along normals).** On a neighborhood $\mathcal{U}_t$ of the normal manifold, $f_t(y) = -\log p_{t,N}(y)$ is $\mu_t$-strongly convex along normal directions:

$$\big(\nabla f_t(y) - \nabla f_t(y^\star)\big)^\top (y - y^\star) \;\geq\; \mu_t \|y - y^\star\|_2^2. \tag{14}$$

Equivalently, $s_N = -\nabla f_t$ is $\mu_t$-strongly anti-monotone:

$$\big(s_N(y) - s_N(y^\star)\big)^\top (y - y^\star) \;\leq\; -\mu_t \|y - y^\star\|_2^2. \tag{15}$$

(Background on monotone operators and strong convexity [7].)

**A2 (Lipschitz score).** On $\mathcal{U}_t$, $\|s_N(y) - s_N(y')\|_2 \leq L_t \|y - y'\|_2$ and $\|\nabla s_N(y)\|_{\mathrm{op}} \leq L_t$.

**A3 (Score approximation).** $S_\theta(\cdot, t) = s_N(\cdot, t) + \xi_t(\cdot)$ with $\|\xi_t\|_\infty \leq \delta_t$ and $\|\nabla \xi_t\|_{\mathrm{op}} \leq \bar{\delta}_t$ on $\mathcal{U}_t$.

**A4 (Reverse integrator).** A one-step scheme with step $h_t > 0$:

$$z_{t+1} = z_t + h_t g_t S_\theta(z_t, t) + \eta_t, \qquad \|\eta_t\| \leq c_t h_t^2, \tag{16}$$

$$h_t g_t L_t \leq \beta_t < 1. \tag{17}$$

(One-step stability/truncation as in standard ODE solvers [15].)

**A5 (Mixture smoothness).** $\gamma_A(\cdot, t)$ and $\Delta s(\cdot, t)$ are locally Lipschitz on $\mathcal{U}_t$.

**A6 (Domain nuisance).** Between source and target, $y_t^{(T)} = y_t^{(S)} + b_D(y_t^{(S)}, t)$ with $\|b_D(y,t) - b_D(y',t)\| \leq \kappa_t \|y - y'\|_2$ and $\|\nabla b_D(y,t)\|_{\mathrm{op}} \leq \lambda_t$.

## S.2 Distribution–residual coupling (what is the anomaly signal?)

**Score mixture decomposition.** By (11), the time-$t$ score decomposes as

$$s(y,t) = \gamma_N(y,t)\, s_N(y,t) + \gamma_A(y,t)\, s_A(y,t), \qquad \text{where } \gamma_N = 1 - \gamma_A, \quad \Delta s = s_A - s_N. \tag{18}$$

**Theorem 0.1** (Directional coupling). *Let $\Delta s_\perp$ be the component of $\Delta s$ orthogonal to $v_t(u)$ in (6). Under A1–A5, there exist constants $c_1 > 0$ and $c_2, c_3 \geq 0$ such that*

$$\mathbb{E}\left\|\bar{M}_t(P)\right\|_2 \;\geq\; c_1\, \mathbb{E}\,\overline{\gamma_A \|\Delta s_\perp\|} - c_2\, \mathbb{E}\left\|\bar{R}_t(P)\right\|_2 - c_3\, \delta_t. \tag{19}$$

**Theorem 0.2** (Path–amplitude coupling). *Along the reverse flow (3), there exists a small remainder $\zeta_t(u)$ with $\|\zeta_t(u)\| \leq C\delta_t$ such that*

$$\partial_t Q_t(u) = g(t)\, R_t(u) + \zeta_t(u),$$
$$\sum_{t=1}^{T} \left\|\bar{Q}_t(P)\right\|_2^2 \geq c_5 \sum_{t=1}^{T} \left\|\bar{R}_t(P)\right\|_2^2 - c_6 \sum_{t=1}^{T} \delta_t^2. \tag{20}$$

**Theorem 0.3** (Temporal stationarity break). *If $\gamma_A(\cdot, t)$ is time-smooth with bounded derivative and $\Delta s(\cdot, t)$ is time-smooth, then there exist $\{a_t > 0\}$ and $b \geq 0$ such that*

$$\mathbb{E}\,\mathrm{NS}_R(P) \;\geq\; \sum_{t=1}^{T-1} a_t\, \mathbb{E}\,\overline{\gamma_A^2 \|\Delta s\|_2^2}\, w_t - b\left(\sum_{t=1}^{T} \mathbb{E}\left\|\bar{R}_t(P)\right\|_2^2 w_t\right)^{-1/2}. \tag{21}$$

*An analogous inequality holds for* $\mathrm{NS}_M(P)$.

**Theorem 0.4** (Curvature lower bound). *Under A1–A3, for any $u \in \Omega$,*

$$\big\|R_t(u)\big\| \;\geq\; \mu_t \,\big\|y_t(u) - P_t(y_t)(u)\big\|_2 \;-\; \delta_t, \tag{22}$$

*so $E_R(P)$ lower-bounds the off-manifold discrepancy up to $\delta_t$ and patch averaging.*

**Theorem 0.5** (Multi-view fusion improves power). *Let $Z(P) = [E_R, E_M, E_Q, \mathrm{NS}_R, \mathrm{NS}_M, \mathrm{NS}_Q, \mathrm{DV}]^\top$. If under $H_0$ (normal) $Z$ has mean $\mu_0$ and covariance $\Sigma_0$, and under $H_1$ (anomaly) mean $\mu_1$, the optimal linear statistic attains*

$$\mathrm{SNR}^\star = \sqrt{(\mu_1-\mu_0)^\top \Sigma_0^{-1}(\mu_1-\mu_0)} \;\geq\; \max_i \frac{|(\mu_1-\mu_0)_i|}{\sqrt{(\Sigma_0)_{ii}}}, \tag{23}$$

*strictly dominating any single component unless $\Sigma_0$ is diagonal and only one component shifts.*

**Interpretation.** Relations (19)–(23) characterize, *before invoking dynamics*, which residual parts are attributable to distributional discrepancy (via $\gamma_A$ and $\Delta s$), motivating $(R, M, Q)$ and their temporal statistics as anomaly-sensitive observables.

## S.3 REF as a spatio–temporal field (tooling the coupling)

**Field definition.** The Residual–Evolution Field (REF) is

$$\mathcal{F}(u,t) = \big(R_t(u),\, M_t(u),\, Q_t(u)\big) \in \mathbb{R}^{3C}. \tag{24}$$

Patch-time statistics (8)–(10) summarize amplitude, directional, and pathwise behaviors with temporal weighting $w_t$ and ridge $\varepsilon_{\mathrm{rid}}$.

**Consistency with the coupling.** By (19)–(22) and (20), $(R, M, Q)$ respectively estimate (i) off-manifold amplitude, (ii) orthogonal anomalous direction, and (iii) accumulated drift; (21) justifies the stationarity test; (23) supports multi-view fusion.

## S.4 Normal-dynamics absorption (why benign residuals vanish?)

**Projection contraction on normals.**

**Lemma 1** (Same-time projection contracts). *Under A1–A3, let $y^\star$ satisfy $s_N(y^\star, t) = 0$. Then*

$$\|P_t(y) - y^\star\|_2^2 \;\leq\; \Big(1 - 2\mu_t\sigma_t^2 + (\tilde{L}_t\sigma_t^2)^2\Big)\|y - y^\star\|_2^2 + C_t\sigma_t^4, \tag{25}$$

*where $\tilde{L}_t = L_t + \bar{\delta}_t$ and $C_t = 2\sigma_t^2\|y - y^\star\|_2\,\delta_t + \sigma_t^4\delta_t^2$.*

**Flow contraction (continuous and discrete).**

**Lemma 2** (Continuous). *Under A1–A3, for solutions of (3),*

$$\begin{aligned}\frac{d}{dt}\|z^{(1)} - z^{(2)}\|_2^2 \leq &-2g(t)\mu_t\|z^{(1)} - z^{(2)}\|_2^2 \\ &+ 2g(t)\delta_t\|z^{(1)} - z^{(2)}\|_2.\end{aligned} \tag{26}$$

**Lemma 3** (Discrete). *Under A1–A4,*

$$\begin{aligned}\|z_{t+1}^{(1)} - z_{t+1}^{(2)}\|_2^2 \leq \rho_t^2\,\|z_t^{(1)} - z_t^{(2)}\|_2^2 + \tilde{c}_t h_t^2, \\ \rho_t^2 = 1 - 2\mu_t h_t g_t + \beta_t^2.\end{aligned} \tag{27}$$

*with $\tilde{c}_t$ depending on $\delta_t, \bar{\delta}_t, c_t$.*

(Contraction uses monotonicity along normals [1, 7] and one-step stability [15].)

**Absorption and non-accumulation.** Let the displacement to the projection be

$$D_t(u) = \big\|y_t(u) - P_t(y_t)(u)\big\|_2. \tag{28}$$

By the mean-value form of (5), there is $\tilde{L}_t$ such that

$$\|R_t(u)\|_2 \;\leq\; \tilde{L}_t\, D_t(u) + \delta_t. \tag{29}$$

**Theorem 0.6** (Absorption on normals). *Under A1–A4, there exist $\kappa_t, \kappa_t' \in (0,1)$ and $B_t, B_t' \geq 0$ so that, for any patch $P$,*

$$\mathbb{E}\, D_{t+1}(P) \leq \kappa_t\, \mathbb{E}\, D_t(P) + B_t, \tag{30}$$

$$\mathbb{E}\,\big\|\bar{R}_{t+1}(P)\big\|_2^2 \leq \kappa_t'\,\mathbb{E}\,\big\|\bar{R}_t(P)\big\|_2^2 + B_t', \tag{31}$$

*and hence*

$$\sum_{t=1}^{T}\mathbb{E}\,\big\|\bar{R}_t(P)\big\|_2^2 \;\leq\; \frac{1}{1-\bar{\kappa}'}\Big(\mathbb{E}\,\big\|\bar{R}_1(P)\big\|_2^2 + \sum_{t=1}^{T} B_t'\Big). \tag{32}$$

*Here $\bar{\kappa}' = \sup_t \kappa_t'$ on the mid/late time range where the one-step stability holds.*

**Corollary 0.7** (Small stationarity indices on normals). *Under Theorem 0.6 and A4,*

$$\begin{aligned}\mathbb{E}\,\mathrm{NS}_R(P) &\leq C \cdot \frac{\sum_{t=1}^{T-1}\mathbb{E}\,\big\|\bar{R}_{t+1}(P) - \bar{R}_t(P)\big\|_2^2\, w_t}{\sum_{t=1}^{T}\mathbb{E}\,\big\|\bar{R}_t(P)\big\|_2^2\, w_t + \varepsilon_{\mathrm{rid}}} \\ &\leq C' \cdot \max_t(1 - \kappa_t').\end{aligned} \tag{33}$$

*with constants $C, C'$ independent of $P$. Identical statements hold for $\mathrm{NS}_M$ and $\mathrm{NS}_Q$.*

## S.5 Anomalous persistence and stationarity break

**Directional persistence.** By (19), nonzero responsibility $\gamma_A$ and orthogonal gap $\Delta s_\perp$ impose a positive lower bound on $\|\bar{M}_t(P)\|_2$ up to $\bar{R}_t$ and approximation terms.

**Accumulated drift persists.** By (20), sustained amplitude induces sustained accumulated drift.

**Stationarity break under time-varying responsibility.** By (21), $\mathrm{NS}_R$ (and $\mathrm{NS}_M$) increases with the temporal variability of $\gamma_A \Delta s$.

## S.6 Discrete-time stability and bounds on stationarity

**Proposition 0.8** (Discrete Grönwall absorption)**.** *Under Lemma 3, if $\rho_t \le \bar{\rho} < 1$ and $\sum_t h_t^2 < \infty$, then*

$$\sum_{t=1}^{T-1} \left\|\bar{R}_{t+1}(P) - \bar{R}_t(P)\right\|_2^2 \;\le\; C\left\|\bar{R}_1(P)\right\|_2^2 + C' \sum_{t=1}^{T} h_t^2, \tag{34}$$

*so $\mathrm{NS}_R(P) = O(1-\bar{\rho})$ uniformly in $P$ for some $\bar{\rho} \in (0,1)$. Identical statements hold for $M, Q$. See discrete Grönwall–type bounds in [14].*

## S.7 REF contracts domain discrepancy; CFA aligns; target risk bound

**REF as a contraction in feature space.** Define the time-$t$ operator

$$\begin{aligned}\mathcal{F}_t(y) &= S_\theta(y,t) - S_\theta\big(y + \sigma_t^2 S_\theta(y,t), t\big) \\ &= \Big[\int_0^1 \nabla_y S_\theta(y + \tau\sigma_t^2 S_\theta, t)\, d\tau\Big] \\ &\quad \big(-\sigma_t^2 S_\theta(y,t)\big).\end{aligned} \tag{35}$$

If $L_{\mathcal{F},t}$ bounds the Lipschitz constant of $\mathcal{F}_t$, then

$$\begin{aligned}W_1\big(\mathcal{F}_t\#P_S,\ \mathcal{F}_t\#P_T\big) \le L_{\mathcal{F},t}\, W_1(P_S,P_T), \quad \text{and} \\ W_1\big(\Phi_{\mathrm{REF}}\#P_S,\ \Phi_{\mathrm{REF}}\#P_T\big) \le \Big(\sum_t w_t L_{\mathcal{F},t}\Big) W_1(P_S,P_T).\end{aligned} \tag{36}$$

Here $\Phi_{\mathrm{REF}}$ denotes the time-aggregated and patch-averaged REF mapping. These follow from the non-expansivity of $W_1$ under Lipschitz maps [12].

**CFA: time, second-order, and directional alignment.** Let $\psi$ be a monotone time reparameterization minimizing

$$\sum_{t=1}^{T} \left\|\mathbb{E}_{P_T}\bar{R}_t - \mathbb{E}_{P_S}\bar{R}_{\psi(t)}\right\|_2, \tag{37}$$

where $\mathbb{E}_{P_D}$ denotes expectation over patches extracted from domain $D \in \{S,T\}$. For stacked REF features $Z_S, Z_T$ (each formed by collecting the per-patch vectors $Z(P) = [E_R, E_M, E_Q, \mathrm{NS}_R, \mathrm{NS}_M, \mathrm{NS}_Q, \mathrm{DV}]^\top$ over source/target), whiten/recolor by

$$Z_T' = \Sigma_S^{1/2}\,\Sigma_T^{-1/2}\,(Z_T - m_T), \tag{38}$$

where $m_T = \mathbb{E}Z_T$, $\Sigma_T = \mathrm{Cov}(Z_T)$, and $\Sigma_S = \mathrm{Cov}(Z_S)$; here $A^{1/2}$ denotes the principal matrix square root (and $A^{-1/2}$ its inverse). Then $\Sigma_{Z_T'} = \Sigma_{Z_S}$ and $\|\mathbb{E}Z_T' - \mathbb{E}Z_S\|_2 \le \Delta_{\mathrm{time}}$. Let $U_S, U_T$ be the top-$r$ principal subspaces of stacked $\{\bar{M}_t\}_t$ (for a fixed subspace dimension $r \in \mathbb{N}$); an orthogonal Procrustes rotation $R^\star$ minimizes $\|U_T R - U_S\|_F$ and reduces directional mismatch. These alignment components follow [3, 8, 11].

**Target risk bound.** For a hypothesis class $\mathcal{H}$ on aligned features $\Phi = \Phi_{\mathrm{CFA}} \circ \Phi_{\mathrm{REF}}$ and bounded loss,

$$\varepsilon_T(h\circ\Phi) \;\le\; \varepsilon_S(h\circ\Phi) \;+\; d_{\mathcal{H}\Delta\mathcal{H}}\big(\Phi\#P_S, \Phi\#P_T\big) \;+\; \lambda^\star, \tag{39}$$

where $d_{\mathcal{H}\Delta\mathcal{H}}$ is the $\mathcal{H}\Delta\mathcal{H}$-divergence and $\lambda^\star = \inf_{h\in\mathcal{H}}\big(\varepsilon_S(h\circ\Phi) + \varepsilon_T(h\circ\Phi)\big)$. Moreover, under A2–A6,

$$\begin{aligned}d_{\mathcal{H}\Delta\mathcal{H}}\big(\Phi\#P_S, \Phi\#P_T\big) \le c_7\Big(\sum_t w_t L_{\mathcal{F},t}\Big) W_1(P_S,P_T) \\ + c_8\,\Delta_{\mathrm{time}},\end{aligned} \tag{40}$$

so if $\big(\sum_t w_t L_{\mathcal{F},t}\big) W_1(P_S,P_T) + \Delta_{\mathrm{time}} \le \delta$ and a source-selected $h^\star$ satisfies $\varepsilon_S(h^\star\circ\Phi) \le \tau$, then for some constant $C_7 > 0$,

$$\varepsilon_T(h^\star\circ\Phi) \;\le\; \tau \;+\; C_7\,\delta \;+\; \lambda^\star. \tag{41}$$

The structure of (39) follows the classic $\mathcal{H}\Delta\mathcal{H}$-divergence bound [2] adapted here to field-space features.

## S.8 Experiments

### Dataset setup

**MVTec AD.** The MVTec Anomaly Detection (MVTec AD) dataset serves as a comprehensive real-world benchmark specifically designed for unsupervised anomaly detection in industrial inspection scenarios. It comprises over 5,000 high-resolution images spanning 15 distinct categories, including five texture categories and ten object categories. Within the source domain setting, full access to annotation information is available. For the target domain, the training set is entirely unlabeled and potentially anomaly-contaminated. Complete annotation information is accessed only during evaluation for quantitative assessment.

Table 1. AUPRC (%) comparison on 9 benchmark datasets. Best per column in **bold**, second best is underlined.

| Methods | Source Domain | MVtec (Bottle) | | | VisA (candle) | | | DAGM (Class 2) | | | Ave. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Target Domain | Cable | Capsule | Hazelnut | Macaroni1 | Macaroni2 | Pcb2 | Class 1 | Class 3 | Class 6 | |
| Unsupervised Methods | General-AD | 51.00 | 77.40 | 55.40 | 25.80 | 14.90 | 27.40 | 26.60 | 32.70 | 81.60 | 43.64 |
| | GLASS | 31.30 | 46.60 | 37.10 | 80.06 | 24.81 | 54.77 | 80.19 | 53.96 | 34.14 | 49.21 |
| | DDPM | 49.63 | 47.43 | 52.34 | 66.86 | 55.17 | 51.92 | 62.09 | 54.61 | 58.11 | 55.35 |
| | DDAD | 54.83 | 55.39 | 58.18 | 76.62 | 70.92 | 57.63 | 69.30 | 71.25 | 63.00 | 64.12 |
| Domain Adaptation | JWO | 42.47 | 47.76 | 48.70 | 67.26 | 70.42 | 68.08 | 22.95 | 23.61 | 82.47 | 52.64 |
| | PWAN | 58.51 | 69.81 | 57.51 | 59.23 | 60.97 | 69.63 | 26.60 | 29.10 | 23.34 | 50.52 |
| | MLWE | 46.82 | 70.99 | 58.61 | 74.37 | 75.89 | 89.30 | 28.76 | 29.10 | 32.37 | 56.25 |
| | CMKD | 65.47 | 54.73 | **66.64** | 72.70 | 60.27 | 87.78 | 17.51 | 55.15 | 85.93 | 62.91 |
| | UniNet | 53.01 | 51.54 | 30.19 | 39.99 | 41.67 | 83.31 | 23.18 | 26.87 | 35.11 | 42.76 |
| | ANC | 43.06 | 55.97 | 19.95 | 41.55 | 52.61 | 42.18 | 24.02 | 23.79 | 32.07 | 37.24 |
| | FFTAT | 59.78 | 59.05 | 65.46 | 82.21 | 87.66 | 84.29 | 22.98 | 41.42 | 70.45 | 63.70 |
| Domain Generalization | GGA | 44.67 | 32.32 | 22.87 | 65.37 | 57.68 | 69.59 | 25.28 | 26.65 | 45.80 | 43.36 |
| | BDC | 38.80 | 47.41 | 26.12 | 43.29 | 46.59 | 47.06 | 26.88 | 23.38 | 52.29 | 39.09 |
| | DDDG | 58.40 | 49.18 | 46.08 | 68.18 | 63.03 | 87.15 | 21.56 | 24.90 | 26.58 | 49.45 |
| | PMGDG | 54.20 | 52.72 | 33.64 | 73.72 | 76.67 | 59.42 | 33.15 | 23.78 | 25.42 | 48.08 |
| | DKGPL | 58.83 | 53.36 | 58.31 | 73.73 | 67.19 | 61.08 | 33.88 | 28.98 | 45.76 | 53.46 |
| Proposed | REF+CFA | **80.01** | **78.89** | 64.80 | **100.00** | **97.00** | **95.94** | **100.00** | **100.00** | **100.00** | **90.74** |

**VisA.** The VisA dataset is a large-scale benchmark for visual anomaly detection, designed to present more complex and realistic industrial scenarios. It contains 10,821 images (9,621 normal and 1,200 anomalous) across 12 object categories, with challenges such as complex structures, multiple instances per image, and diverse low-level defects. In this work, only sample-level labels are considered. Under the cross-domain setting, the source domain provides sample-level supervision for training, whereas the target-domain training partition is entirely unlabeled and potentially anomaly-contaminated. Target-domain label information is accessed only during evaluation for rigorous quantitative analysis.

**DAGM 2007.** The DAGM 2007 dataset is a synthetic benchmark for industrial optical inspection, designed to detect diverse defects on statistically textured backgrounds. The data are artificially generated and divided into ten subsets, including six development sets and four competition sets. Each development set contains 1000 non-defective images and 150 defective images. In this work, only sample-level labels are considered. Under the cross-domain setting, the source domain is fully labeled at the sample level, while the target-domain training set is entirely unlabeled and potentially anomaly-contaminated. Target-domain label information is accessed only during evaluation for quantitative assessment.

## comparison with SOTA

Tables 1 and 2 present the AUPRC and Rec@K scores, respectively, further validating the robustness of our method.

**Performance on AUPRC** For AUPRC, the proposed REF+CFA achieves a remarkable average score of **90.74%**, surpassing the strongest baseline (*DDAD*) by a substantial margin of 26.62 percentage points (64.12%). On the challenging VisA benchmarks, REF+CFA demonstrates exceptional precision, reaching **100.00%** on Macaroni1 and **97.00%** on Macaroni2, outperforming the runner-up by up to 17.79 points (e.g., on Macaroni1). This advantage is even more pronounced on DAGM, where REF+CFA attains perfect AUPRC (100.00%) across all target classes (Class 1, Class 3, and Class 6), exceeding the best prior methods by 19.81, 28.75, and 14.07 points, respectively.

**Performance on Rec@K** In terms of Rec@K, REF+CFA exhibits comprehensive superiority, achieving the best performance across all nine source–target transfers with an average of **88.34%**. This corresponds to a substantial gain of 26.97 points over the strongest competing method (*DDAD*, **61.37%**). Notably, REF+CFA attains perfect recall (**100.00%**) on four scenarios (VisA Macaroni1 and all DAGM classes) and consistently leads on the MVTec benchmarks (e.g., 70.63% vs. 65.22% on Cable). These results confirm that the proposed framework effectively aligns anomaly features across domains, thereby improving the retrieval of anomalous samples under cross-domain shifts.

## Qualitative Results

To provide a more intuitive understanding of the proposed REF+CFA framework's anomaly localization capabilities, we present a representative qualitative comparison in Figure 1. The visualization includes the original input im-

Table 2. Rec@$K$ (%) comparison on 9 benchmark datasets. Best per column in **bold**, second best is underlined.

| Methods | Source Domain | MVtec (Bottle) | | | VisA (candle) | | | DAGM (Class 2) | | | Ave. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Target Domain | Cable | Capsule | Hazelnut | Macaroni1 | Macaroni2 | Pcb2 | Class 1 | Class 3 | Class 6 | |
| Unsupervised Methods | General-AD | 48.90 | 66.10 | 52.90 | 33.34 | 18.17 | 31.79 | 28.40 | 35.30 | 73.30 | 43.13 |
| | GLASS | 35.71 | 46.67 | 42.42 | 72.91 | 33.19 | 53.80 | 43.61 | 47.72 | 35.33 | 45.71 |
| | DDPM | 47.12 | 41.41 | 51.81 | 63.85 | 54.71 | 48.81 | 49.79 | 51.81 | 49.13 | 50.94 |
| | DDAD | 50.48 | 53.41 | 51.33 | 78.41 | 59.79 | 54.52 | 64.22 | 74.21 | 65.92 | 61.37 |
| Domain Adaptation | JWO | 65.22 | 67.27 | 65.71 | 62.12 | 62.87 | 64.71 | 20.35 | 23.68 | 73.68 | 56.18 |
| | PWAN | 50.00 | 56.36 | 60.29 | 57.65 | 61.66 | 63.49 | 24.91 | 32.89 | 25.00 | 48.03 |
| | MLWE | 41.30 | 63.64 | 65.71 | 67.77 | 63.83 | 75.29 | 27.03 | 22.91 | 32.43 | 51.10 |
| | CMKD | 54.55 | 51.02 | 60.61 | 66.27 | 58.88 | 75.97 | 16.19 | 46.30 | 76.35 | 56.24 |
| | UniNet | 47.83 | 49.09 | 33.49 | 41.18 | 39.87 | 74.92 | 23.68 | 28.95 | 31.58 | 41.18 |
| | ANC | 45.65 | 56.36 | 17.14 | 38.88 | 56.52 | 34.65 | 24.95 | 22.37 | 30.26 | 36.31 |
| | FFTAT | 53.33 | 56.00 | 58.66 | 73.56 | 78.12 | 78.32 | 20.51 | 43.59 | 64.10 | 58.47 |
| Domain Generalization | GGA | 39.13 | 30.28 | 22.86 | 60.19 | 58.72 | 59.09 | 23.68 | 26.32 | 50.00 | 41.14 |
| | BDC | 19.57 | 16.36 | 28.57 | 35.42 | 48.99 | 43.59 | 24.71 | 18.11 | 44.34 | 31.07 |
| | DDDG | 50.00 | 47.27 | 57.14 | 63.78 | 60.63 | 74.59 | 23.68 | 23.68 | 24.32 | 47.23 |
| | PMGDG | 45.65 | 56.36 | 28.57 | 65.91 | 63.83 | 51.57 | 35.53 | 23.68 | 24.52 | 43.96 |
| | DKGPL | 63.04 | 63.64 | 68.57 | 68.24 | 66.18 | 52.58 | 33.21 | 31.58 | 43.42 | 54.50 |
| Proposed | REF+CFA | **70.63** | **69.39** | **70.56** | **100.00** | **91.78** | **92.69** | **100.00** | **100.00** | **100.00** | **88.34** |

age, the corresponding ground-truth (GT) anomaly mask, the qualitative anomaly detection results produced by our method, and those from the GeneralAD baseline. As observed, our method aligns more closely with the ground-truth mask, demonstrating superior precision in capturing the anomalous regions while effectively suppressing background noise in the target domain.

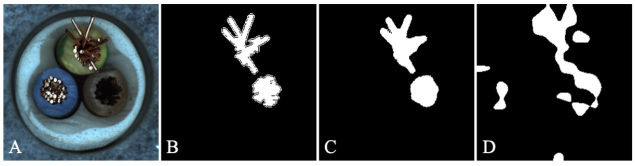


Figure 1. Qualitative comparison of anomaly localization. A: Input image; B: Ground-truth (GT) anomaly mask; C: Ours (REF+CFA); D: GeneralAD baseline.

## Runtime and Efficiency Analysis

In practical industrial deployments, the computational overhead of diffusion-based methods is a critical consideration. While our theoretical analysis leverages reverse diffusion dynamics, the REF framework does not require a full reverse trajectory in practice. Instead, the sampling step count, denoted as $K$, explicitly controls the efficiency–accuracy trade-off.

Table 3 illustrates this trade-off using the Hazelnut category as a representative example. We observe that REF achieves highly competitive performance even at a significantly reduced step count. Specifically, increasing $K$ from 40 to 160 yields diminishing returns in AUROC while substantially increasing the training time. This demonstrates that the proposed framework can be flexibly adjusted to meet specific latency and resource constraints without strictly adhering to the complete diffusion trajectory.

Table 3. Runtime–performance trade-off with respect to the sampling step count $K$ (evaluated on the Hazelnut category).

| Step Count | $K$=10 | $K$=20 | $K$=40 | $K$=80 | $K$=160 |
|---|---|---|---|---|---|
| AUROC (%) | 78.42 | 88.85 | 91.66 | 91.38 | 90.91 |
| Training Time | 0.31h | 0.54h | 0.85h | 1.55h | 3.28h |

## References


[1] Heinz H. Bauschke and Patrick L. Combettes. Convex analysis and monotone operator theory in hilbert spaces. In *CMS Books in Mathematics*, 2011. 3

[2] 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. 4

[3] Marco Cuturi and Mathieu Blondel. Soft-dtw: a differentiable loss function for time-series. In *International conference on machine learning*, pages 894–903. PMLR, 2017. 4

[4] Bradley Efron. Tweedie's formula and selection bias. *Journal of the American Statistical Association*, 106(496):1602–1614, 2011. 1

[5] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 1

[6] Aapo Hyvärinen. Estimation of non-normalized statistical models by score matching. *J. Mach. Learn. Res.*, 6:695–709, 2005. 1

[7] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*. Springer Science & Business Media, 2013. 2, 3

[8] Peter H Schönemann. A generalized solution of the orthogonal procrustes problem. *Psychometrika*, 31(1):1–10, 1966. 4

[9] Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. pmlr, 2015. 1

[10] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *9th International Conference on Learning Representations, ICLR 2021, Virtual Event, Austria, May 3-7, 2021*. OpenReview.net, 2021. 1

[11] Baochen Sun and Kate Saenko. Deep coral: Correlation alignment for deep domain adaptation. In *Computer Vision – ECCV 2016 Workshops*, pages 443–450, Cham, 2016. Springer International Publishing. 4

[12] Cédric Villani. Optimal transport: Old and new. 2008. 2, 4

[13] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011. 1

[14] Wolfgang Walter. *Differential and integral inequalities*. Springer Science & Business Media, 2012. 4

[15] Gerhard Wanner and Ernst Hairer. *Solving ordinary differential equations II*. Springer Berlin Heidelberg New York, 1996. 2, 3