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.

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

Abstract

Label-free image anomaly detection is difficult because anomalies must be separated from intra-normal variability. Diffusion models learn a manifold for normal data, and, under the common assumption that off-manifold anomalies are harder to generate and yield larger prediction errors, many methods build detectors from prediction residuals; yet reverse-process stochasticity and complex but normal structure also produce large residuals, so magnitude alone is non-diagnostic. To clarify what is recoverable from such noisy residuals, the theory examines how residual signals propagate through later reverse steps, showing that variability consistent with normal statistics is gradually absorbed toward stationarity, whereas anomalous regions retain an additional non-stationary signal that persists. Building on this insight, the Residual–Evolution Field (REF) isolates this persistent signal, with labeled source data calibrating the extractor and Cross-domain Field Alignment (CFA) transferring it to unlabeled targets. A theoretical framework with formal analysis is established, and experiments across multiple benchmarks under substantial domain shifts demonstrate state-of-the-art performance, improving over strong baselines by 2.01–14 percentage points (pp).

1. Introduction

Label-free image anomaly detection is challenging because anomalies are hard to distinguish from the intrinsic variability among normal samples. This ambiguity is amplified in realistic data, where training sets often contain both normal and anomalous images without labels. A detector must separate within-normal variation on the normal manifold from true departures. Conventional unsupervised models often fail to isolate such distributional departures, and transfer methods remain reliable only under small shifts, offering limited domain-agnostic guidance in practice.

*Corresponding author: beiyj@zju.edu.cn

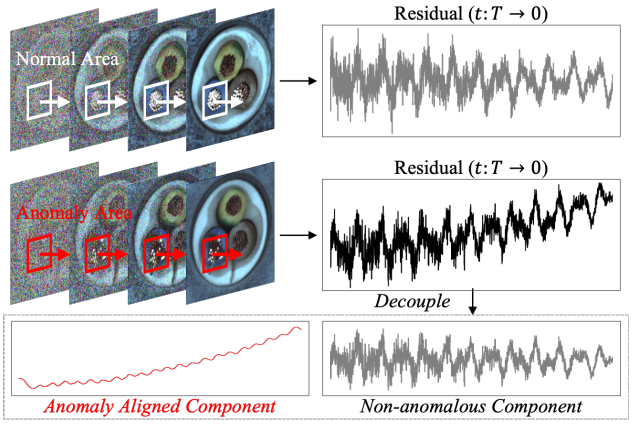


Figure 1. **Motivating schematic for REF.** Residuals in both normal and anomalous regions exhibit strong randomness. In anomalous regions, an additional anomaly-aligned component, explainable by the distributional difference between normal and anomaly, forms a coherent non-stationary trend (red). REF is designed to isolate this hidden component from noisy residuals.

Diffusion models learn a robust normal manifold that explains such within-normal variability. Many diffusion-based approaches therefore learn detectors from *prediction residuals*. However, stochasticity in reverse diffusion and complex local structure can also enlarge residuals, injecting substantial noise into the learned representation and weakening cross-domain generalization. These limitations motivate extracting the anomaly-aligned component from residuals rather than relying on residual magnitude or raw residual features.

To extract the anomaly-aligned component from diffusion residuals, residual-based statistics are designed to quantify how departures from the learned normal distribution influence residual evolution across time and space, providing an interpretable coupling between distributional departure and residual behaviour. Building on this foundation, the analysis examines how residual signals at a given step influence subsequent reverse transitions. It shows that resid-

ual variability consistent with intra-normal statistics is progressively absorbed and moves toward stationarity, whereas residuals from anomalous regions carry an additional non-stationary signal that persists rather than being absorbed. These results motivate testing for a hidden non-stationary component via patch-level energy and stationarity indices, with full derivations and bounds in Supplementary Material (Supp.) Section S.4–S.5.

Building on these results, the Residual–Evolution Field (REF) is introduced to detect the non-stationary, anomaly-aligned signal hidden in diffusion residuals, as illustrated in Fig. 1. Direct training on unlabeled data is unreliable because this signal is buried under reverse-diffusion stochasticity and intra-normal variability. Labeled *source* data are therefore used to train and calibrate the REF extractor, and a Cross-domain Field Alignment (CFA) module aligns the field space so the detector transfers to unlabeled *target* domains. Operating in the aligned field space reduces domain discrepancy and admits a target-risk bound. The full derivation appears in Supp. S.7.

This work develops the REF to isolate the anomaly-related signal embedded in diffusion residuals, together with CFA for label-free transfer across domains. A theoretical framework with formal support is established, and extensive experiments under substantial domain shifts demonstrate consistent state-of-the-art performance, improving over strong baselines by 2.01–14 pp.

2. Related Work

Prior work falls into three categories: unsupervised modeling of normality, supervised anomaly detection, and cross-domain transfer learning.

2.1. Unsupervised anomaly detection

Unsupervised methods learn normality and identify anomalies by deviations from it. These methods can be grouped into four categories.

Feature matching and distillation. These methods characterize normality in pre-trained feature space and score anomalies by deviations from the learned feature patterns. PaDiM models per-location features as multivariate Gaussians and uses Mahalanobis scoring [10]. SPADE builds dense multi-scale correspondences to a nominal gallery for pixel-level anomaly maps [9]. PatchCore compresses nominal features into a coreset to enable fast kNN retrieval and high-recall detection [29]. STPM trains a student to align teacher feature pyramids so that mismatches indicate anomalies [36]. RD4AD reverses the distillation direction to sharpen residual signals and improve localization [11]. Patch SVDD optimizes a patch-level one-class objective for segmentation [45].

Density estimation and reconstruction. These methods model normal data distributions or reconstruction patterns and treat poorly explained content as anomalous. CFLOW-AD employs conditional normalizing flows with multi-scale decoders for real-time scoring [15]. FastFlow performs 2D flow-based density estimation over features for efficient localization [47]. MemAE adds a prototypical memory to an autoencoder to constrain reconstructions to normal patterns [13]. GANomaly couples adversarial and reconstruction losses to expose feature-level deviations [1]. f-AnoGAN inverts a trained GAN and detects anomalies via residual and latent distances [32].

Diffusion models. These methods characterize normality through denoising dynamics and expose anomalies through abnormal diffusion responses. AnoDDPM reconstructs by DDPM and aggregates multi-scale trajectory errors for detection and localization [42]. Denoising-diffusion OOD computes score norms or denoising consistency as an anomaly cue [14]. DiffAD leverages latent diffusion with noise-conditioned embeddings to improve industrial anomaly localization [54]. Unsupervised industrial AD with diffusion models evaluates diffusion-based reconstruction for real-world inspection [43]. Unsupervised OOD with diffusion uses diffusion manifolds for target–inlier separation [24]. DiAD formulates a diffusion-based framework for multi-class anomaly detection with pixel- and feature-space cues [16].

Vision–language models. These methods leverage semantic priors from image–text pretraining to improve anomaly detection under limited supervision. WinCLIP uses windowed CLIP features and prompt ensembles for classification and segmentation [19]. PromptAD learns task-specific prompts from normal samples to enhance few-shot performance [21]. AnomalyCLIP formulates object-agnostic prompt learning for zero-shot industrial AD [57].

Limitations. Unsupervised methods mainly constrain normal samples to conform to learned normal patterns, but they do not explicitly enforce separation between normal and anomalous samples. This makes it difficult to distinguish benign perturbations of normal samples from subtle anomalous deviations, particularly when the anomaly signal is weak and easily mixed with intra-normal variability.

2.2. Supervised anomaly detection

Supervised anomaly detection methods rely on anomaly-related supervision to learn discriminative anomaly cues. In practice, these methods mainly fall into two types.

Synthetic supervision. These methods create supervision without real anomaly labels by simulating synthetic anomalies. CutPaste synthesizes paste-based perturbations to train a anomaly classifier and segmenter [20]. DRAEM combines a reconstruction pathway with a discriminative branch trained on simulated masks [49]. DeSTSeg applies segmentation-guided denoising in a student–teacher framework for fine-grained maps [55]. RealNet generates realis-

tic synthetic anomalies with feature selection to reduce bias [56]. Outlier Exposure regularizes models with auxiliary outliers to improve open-set robustness [17].

Real-anomaly supervision. These methods learn decision boundaries directly from annotated anomalies or limited labels. SimpleNet uses a lightweight design with multi-scale cues for pixel-level detection [25]. Open-set supervised AD explicitly separates known and unknown regions during training [12]. Boundary-guided contrastive learning exploits explicit boundaries to refine localization [44]. Deep SAD optimizes a hypersphere-based objective with limited labels [30]. DevNet employs a deviation-based loss to model anomaly intensity [28].

Limitations. Supervised pipelines tend to overfit the anomaly modes observed during training and degrade on rare or unseen anomalies. Pixel-level labels are expensive. Synthetic anomalies reduce annotation cost but introduce artifacts and domain gaps that bias the decision boundary.

2.3. Transfer learning for anomaly detection

Studies on transfer learning for anomaly detection are still limited. Methods adaptable to cross-domain anomaly detection are summarized into two main categories.

Domain adaptation. These methods use target-domain information to align shared structure across domains. IWAN estimates instance-level importance to focus alignment on transferable samples [53]. PADA introduces class-level weights to suppress source-only categories [5]. ETN measures example transferability to filter harmful supervision [6]. OSBP adds an explicit unknown class to enable open-set separation [31]. Universal DA discovers shared and unknown classes for a unified treatment [46].

Domain Generalization. In this paper, this category refers to methods that transfer a source-trained hypothesis without requiring target-domain labels. SHOT transfers the classifier and maximizes information on target features [23]. TENT minimizes prediction entropy mainly via batch-normalization updates [35]. EATA regularizes test-time updates to prevent catastrophic forgetting [27]. Test-Time Training adds a self-supervised auxiliary task to improve robustness under shift [34]. AdaBN adapts batch-normalization statistics to the target domain [22].

Limitations. Large domain gaps, especially when normal manifolds differ, can suppress anomaly-specific cues after alignment. Partial and open-set adaptation assume stable label overlap that often does not hold for anomaly detection. Source-free and test-time adaptation reduce reliance on source data but may adapt to spurious correlations and do not explicitly preserve anomaly-sensitive directions.

Summary. Unsupervised methods can model normality, but they do not explicitly constrain anomalous samples to stay far from normal patterns, making it difficult to distinguish benign perturbations from subtle anomalous deviations. Supervised methods improve anomaly discrimination, yet they rely on limited anomalous samples observed during training and therefore tend to overfit seen anomaly patterns. Existing transfer learning methods are also difficult to adapt to anomaly detection scenarios, where target-domain supervision is unavailable and cross-domain heterogeneity further increases the difficulty of preserving anomaly-sensitive information.

3. Method

Overview. This section is organized in three parts. First, a compact but coherent *theory layer* explains why a spatio–temporal *vector field* built from diffusion residuals separates anomaly signal from non–anomalous variability and supports cross–domain reuse. Second, an *architecture layer* presents the end–to–end data flow of **REF+CFA**: from the input image to field construction, statistics, alignment objectives, and decision functions. Third, a brief *summary* includes complexity and deployment notes.

3.1. Part I: Theory

Setup and symbols. An image is $x \in \mathbb{R}^{H\times W\times C}$. The pixel lattice is Ω with locations $u \in \Omega$. Diffusion time is $t \in \{1, 2, \dots, T\}$ with T the data time. The forward perturbation is $y_t = \alpha_t x + \sigma_t \varepsilon$ with $\alpha_t^2 + \sigma_t^2 = 1$ and $\varepsilon \sim \mathcal{N}(0, I)$. A score $S_\theta(y, t) \approx \nabla_y \log p_t(y)$ (time–t density p_t) induces the same–time Tweedie map $P_t(y) = y + \sigma_t^2 S_\theta(y, t)$ and the probability–flow ODE $\dot{z} = g(t)S_\theta(z, t)$ with $g(t) > 0$, whose solution from t to T is $\Phi_{t\to T}(y)$. For a patch $P \subset \Omega$ (center u), define patch–means of a vector field X_t by $\bar{X}_t(P) = |P|^{-1}\sum_{v\in P} X_t(v)$ and time differences by $\Delta_t \bar{X}_t(P) = \bar{X}_{t+1}(P) - \bar{X}_t(P)$. For any per-pixel scalar $q_t(u)$, write $\overline{q_t}(P) = |P|^{-1}\sum_{v\in P} q_t(v)$. Time weights are $w_t \geq 0$ with $\sum_{t=1}^{T} w_t = 1$, and $\epsilon_{\text{rid}} > 0$ (used in denominators below) is a small ridge constant distinct from the forward noise ε. Norm $\|\cdot\|_2$ and inner product $\langle\cdot,\cdot\rangle$ are Euclidean. Indices respect $1, 2, \dots, t, t+1, \dots, T$.

Complete assumptions and analysis appear in the supplementary material (assumptions in Supp. S.1; main results in Supp. S.2–S.7). The following synopsis records only the operational consequences needed in practice:

1) Residual–distribution coupling. Model $p_t = (1 - \pi)p_{t,N} + \pi p_{t,A}$ with responsibility γ_A and scores $s_N = \nabla \log p_{t,N}$, $s_A = \nabla \log p_{t,A}$, $\Delta s = s_A - s_N$. Define the *reference direction* $v_t(u) = S_\theta(P_t(y_t), t)(u)$ and the projection $\Pi_v[w] = \frac{\langle w,v\rangle}{\|v\|_2^2} v$. Introduce the residual primitives at pixel (u, t):

$$R_t(u) = S_\theta(y_t, t)(u) - S_\theta(P_t(y_t), t)(u), \quad (1)$$

$$M_t(u) = S_\theta(y_t, t)(u) - \Pi_{v_t(u)}[S_\theta(y_t, t)(u)], \quad (2)$$

$$Q_t(u) = \Phi_{t\to T}(y_t)(u) - \Phi_{t\to T}(P_t(y_t))(u). \quad (3)$$

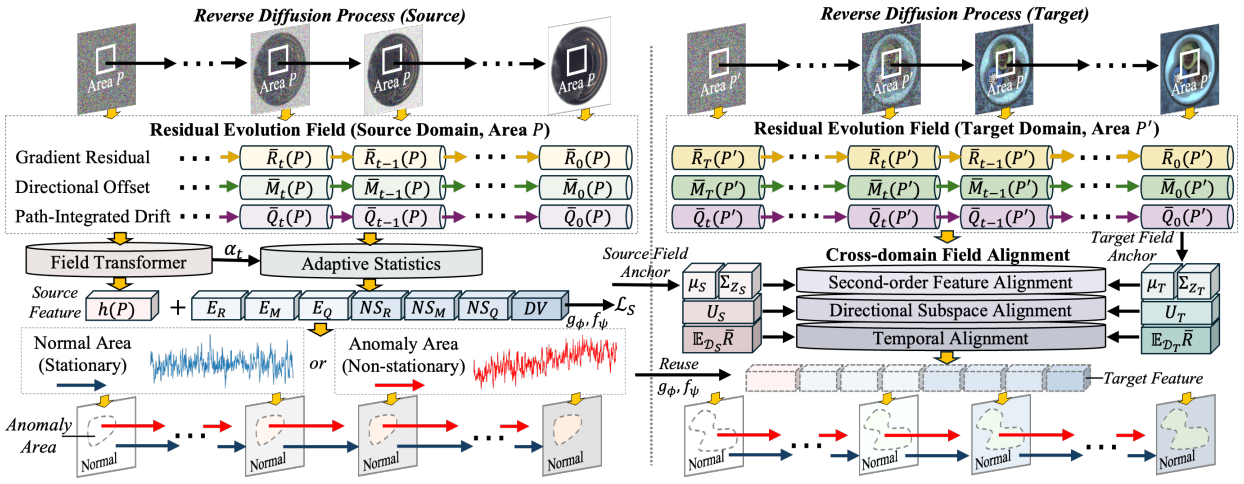


Figure 2. **REF+CFA pipeline.** *Left (source, supervised):* a Residual–Evolution Field is built per patch across diffusion time using three vectors—amplitude residual R, directional offset M, and path-integrated drift Q. A field transformer yields time attention and a compact 7-D REF signature summarizing energy and non-stationarity, where normal areas are more stationary/low-energy and anomalous areas more persistent/non-stationary. *Right (target, unsupervised):* the same REF is computed and *aligned to the source field space* via Cross-domain Field Alignment—temporal alignment, directional subspace alignment, and second-order feature alignment—so that REF response patterns match across domains. This cross-domain consistency is the basis for reusing the source-trained REF without target labels.

Then (Supp. S.2) the following couplings hold: (i) **Amplitude–geometry**: $\|R_t(u)\| \geq \mu_t \|y_t(u) - P_t(y_t)(u)\|_2 - \delta_t$, so E_R lower–bounds off–manifold discrepancy (curvature $\mu_t > 0$, approximation δ_t); (ii) **Directional**: letting $\Delta s_\perp$ be the component of Δs orthogonal to v_t, $\mathbb{E}\|\bar{M}_t(P)\|_2 \geq c_1 \mathbb{E}\overline{\gamma_A \|\Delta s_\perp\|} - c_2 \mathbb{E}\|\bar{R}_t(P)\|_2 - c_3\,\delta_t$; (iii) **Accumulation**: $\partial_t Q_t = g(t) R_t + \zeta_t$ with $\|\zeta_t\| \leq c_4 \delta_t$, hence $\sum_t \|\bar{Q}_t(P)\|_2^2 \geq c_5 \sum_t \|\bar{R}_t(P)\|_2^2 - c_6 \sum_t \delta_t^2$; (iv) **Stationarity break**: temporal variation of $\gamma_A \Delta s$ induces a lower bound on NS_R (and NS_M). A multi–view statistic $Z = [E_R, E_M, E_Q, \mathrm{NS}_R, \mathrm{NS}_M, \mathrm{NS}_Q, \mathrm{DV}]^\top$ can improves SNR over any single component (Supp. S.2, S.6).

2) REF field and stationarity statistics. Define the *Residual–Evolution Field* $\mathcal{F}(u,t) = (R_t(u), M_t(u), Q_t(u))$. On a patch P, use time–weighted energies and stationarity indices

$$E_X(P) = \sum_{t=1}^{T} \|\bar{X}_t(P)\|_2^2\, w_t, \tag{4}$$

$$\mathrm{NS}_X(P) = \frac{\sum_{t=1}^{T-1} \|\bar{X}_{t+1}(P) - \bar{X}_t(P)\|_2^2\, w_t}{\sum_{t=1}^{T} \|\bar{X}_t(P)\|_2^2\, w_t + \epsilon_{\mathrm{rid}}}, \tag{5}$$

$X \in \{R, M, Q\}$, and the directional variability $\mathrm{DV}(P) = \sum_{t=1}^{T-1} \frac{\|\bar{M}_{t+1}(P) - \bar{M}_t(P)\|_2^2}{\|\bar{M}_t(P)\|_2^2 + \epsilon_{\mathrm{rid}}}\, w_t$. Normal regions tend to exhibit small E_X and small NS_X; anomalous regions tend to sustain energy and break stationarity. (Field definition and test: Supp. S.3; couplings supporting them: Supp. S.2.)

3) Contraction and non-accumulation under normal dynamics. Let the *projection displacement* be $D_t(P) = |P|^{-1} \sum_{u \in P} \|y_t(u) - P_t(y_t)(u)\|_2$. Under local strong convexity of the normal log–density along normals and a stable reverse integrator, there exist factors $\kappa_t, \kappa_t' \in (0,1)$ and remainders $B_t, B_t' \geq 0$ (collecting discretization and score–approximation terms) such that

$$\mathbb{E} D_{t+1}(P) \leq \kappa_t\, \mathbb{E} D_t(P) + B_t, \tag{6}$$

$$\mathbb{E}\|\bar{R}_{t+1}(P)\|_2^2 \leq \kappa_t'\, \mathbb{E}\|\bar{R}_t(P)\|_2^2 + B_t'. \tag{7}$$

Consequently $\sum_{t=1}^{T} \mathbb{E}\|\bar{R}_t(P)\|_2^2$ is bounded, and discrete stationarity indices are uniformly small (continuous/discrete derivations and Grönwall bounds in Supp. S.4 and S.6). This formalizes *contraction and non-accumulation* of non-anomalous residuals under the learned normal dynamics.

4) Anomaly-aligned persistence and stationarity break. Under the mixture model above, the directional lower bound and the path–amplitude relation imply that anomaly–induced residuals can *persist* and accumulate into non-vanishing drift:

$$\mathbb{E}\|\bar{M}_t(P)\|_2 \geq c_1\, \mathbb{E}\overline{\gamma_A \|\Delta s_\perp\|} - c_2\, \mathbb{E}\|\bar{R}_t(P)\|_2 - c_3\, \delta_t,$$
$$\sum_t \|\bar{Q}_t(P)\|_2^2 \geq c_5 \sum_t \|\bar{R}_t(P)\|_2^2 - c_6 \sum_t \delta_t^2. \tag{8}$$

Temporal variability of $\gamma_A \Delta s$ further forces a *stationarity break* in R_t and M_t (Supp. S.5).

5) Field alignment for cross-domain reuse and risk bound. Let W_1 be the 1-Wasserstein distance and "#"

the pushforward of measures. For the per-time REF operator $\mathcal{F}_t$, $W_1\big(\mathcal{F}_t\#P_S,\ \mathcal{F}_t\#P_T\big) \le L_{\mathcal{F},t}\, W_1(P_S, P_T)$, $\sum_t w_t L_{\mathcal{F},t} < 1$ at informative times, so REF contracts cross-domain discrepancy in field space (Supp. S.7). CFA then aligns time (monotone reparameterization), second order (whiten–recolor), and direction (orthogonal Procrustes), and yields a field–space transfer bound:

$$\varepsilon_T(h\circ\Psi) \le \varepsilon_S(h\circ\Psi) + d_{\mathcal{H}\Delta\mathcal{H}}\big(\Psi\#P_S, \Psi\#P_T\big) + \lambda^\star, \tag{9}$$

with $\Psi = \Phi_{\mathrm{CFA}}\circ\Phi_{\mathrm{REF}}$ the aligned field mapping, h a detection head, $d_{\mathcal{H}\Delta\mathcal{H}}$ the $\mathcal{H}\Delta\mathcal{H}$ divergence, and $\lambda^\star$ the joint error of the ideal hypothesis. The divergence term is explicitly shrunk by the REF contraction and the CFA terms (Supp. S.7). Thus the *anomaly response pattern* in field space becomes isomorphic across domains, enabling reuse on unlabeled targets.

3.2. Part II: Architecture

Setting. A transfer setup is considered with a *labeled* source domain $\mathcal{D}_S = \{(x_i^S, y_i^S)\}_{i=1}^{N_S}$ and an *unlabeled* target domain $\mathcal{D}_T = \{x_j^T\}_{j=1}^{N_T}$, where images $x \in \mathbb{R}^{H\times W\times C}$. Both domains are *mixed* (normal and anomalous samples co-exist). Labels on $\mathcal{D}_S$ may be image-level $y_i^{\mathrm{img}} \in \{0,1\}$ or patch-level $y(P) \in \{0,1\}$; when only image labels are available, a MIL aggregator $\hat{y}^{\mathrm{img}} = 1 - \prod_P (1-\hat{y}(P))$ is used. The spatial lattice is $\Omega = \{1,\dots,H\}\times\{1,\dots,W\}$; a patch is $P \subset \Omega$ with center u. Time indices are $t \in \{1,2,\dots,T\}$ with T the data time. All $\lambda_\bullet \ge 0$ (e.g., $\lambda_{\mathrm{ns}}, \lambda_{\mathrm{tv}}, \lambda_{\mathrm{stat}}, \lambda_{\mathrm{sp}}, \lambda_{\mathrm{cor}}, \lambda_{\mathrm{sub}}$) and $\eta \ge 0$ are scalar hyper-parameters.

As shown in Figure 2, the method has two components: (i) **REF** construction *per domain*, and (ii) **CFA** alignment to transfer a source-trained detector to $\mathcal{D}_T$ without labels.

A. Source: supervised field detector.

(1) Score and projection. Train a diffusion score network $S_{\theta_S}(y,t) \approx \nabla_y \log p_t^{(S)}(y)$ on $\mathcal{D}_S$ using denoising score matching. Noisy states follow $y_t^S = \alpha_t x^S + \sigma_t \varepsilon$ with schedule $\alpha_t^2 + \sigma_t^2 = 1$ and $\varepsilon \sim \mathcal{N}(0, I)$ (forward noise). The same-time Tweedie projection is $P_t(y) = y + \sigma_t^2 S_{\theta_S}(y,t)$. The reverse (probability–flow) dynamics are $\dot{z} = g(t)\, S_{\theta_S}(z,t)$ with $g(t) > 0$; $\Phi_{t\to T}(y)$ denotes the solution at time T for $z(t) = y$.

(2) REF per patch. For each $P \subset \Omega$, define field components at pixel u and time t (source domain):

$$R_t(u) = S_{\theta_S}(y_t^S, t)(u) - S_{\theta_S}\big(P_t(y_t^S), t\big)(u), \tag{10}$$

$$M_t(u) = S_{\theta_S}(y_t^S, t)(u) - \Pi_{v_t(u)}\big[S_{\theta_S}(y_t^S, t)(u)\big], \tag{11}$$

$$v_t(u) = S_{\theta_S}\big(P_t(y_t^S), t\big)(u), \tag{12}$$

$$Q_t(u) = \Phi_{t\to T}(y_t^S)(u) - \Phi_{t\to T}\big(P_t(y_t^S)\big)(u), \tag{13}$$

where $\Pi_v[w] = \frac{\langle w,v\rangle}{\|v\|_2^2} v$ is the orthogonal projection onto v. Patch averages are $\bar{X}_t(P) = |P|^{-1}\sum_{u\in P} X_t(u)$ for $X \in \{R, M, Q\}$; the operator $\mathrm{vec}(\cdot)$ flattens tensors to vectors.

(3) Field transformer and adaptive time. Form the sequence $\{\mathrm{vec}(\bar{R}_t(P), \bar{M}_t(P), \bar{Q}_t(P))\}_{t=1}^T$ with temporal positional encodings and feed it to a lightweight *Field Transformer* g_ϕ (a generic sequence model). Outputs are (i) *time attention* $\alpha_t(P) = \mathrm{softmax}(\mathrm{logits}_t)$ with $\sum_t \alpha_t(P) = 1$, and (ii) a patch embedding $h(P)$. Using $\alpha_t(P)$, define adaptive statistics for $X \in \{R, M, Q\}$:

$$E_X^{\mathrm{att}}(P) = \sum_{t=1}^{T} \alpha_t(P)\, \big\|\bar{X}_t(P)\big\|_2^2, \tag{14}$$

$$\mathrm{NS}_X^{\mathrm{att}}(P) = \frac{\sum_{t=1}^{T-1} \bar{\alpha}_{t,t+1}(P)\, \big\|\bar{X}_{t+1}(P) - \bar{X}_t(P)\big\|_2^2}{\sum_{t=1}^{T} \alpha_t(P)\, \big\|\bar{X}_t(P)\big\|_2^2 + \varepsilon_{\mathrm{rid}}}, \tag{15}$$

where $\bar{\alpha}_{t,t+1}(P) = \frac{1}{2}\big(\alpha_t(P) + \alpha_{t+1}(P)\big)$ and $\varepsilon_{\mathrm{rid}} > 0$ is a small ridge constant in denominators (distinct from the Gaussian noise ε).

(4) Patch predictor and supervision on $\mathcal{D}_S$. A head f_ψ maps the concatenation:

$$[h(P);\ E_R^{\mathrm{att}}, E_M^{\mathrm{att}}, E_Q^{\mathrm{att}}, \mathrm{NS}_R^{\mathrm{att}}, \mathrm{NS}_M^{\mathrm{att}}, \mathrm{NS}_Q^{\mathrm{att}}, \mathrm{DV}(P)] \mapsto \hat{y}(P) \in [0,1], \tag{16}$$

where the *directional variability* is

$$\mathrm{DV}(P) = \sum_{t=1}^{T-1} \frac{\big\|\bar{M}_{t+1}(P) - \bar{M}_t(P)\big\|_2^2}{\big\|\bar{M}_t(P)\big\|_2^2 + \varepsilon_{\mathrm{rid}}}\, \alpha_t(P). \tag{17}$$

Binary cross-entropy (BCE) is used for patch labels. With only image labels $y^{\mathrm{img}} \in \{0,1\}$, the MIL aggregator $\hat{y}^{\mathrm{img}} = 1 - \prod_P (1-\hat{y}(P))$ is trained with $\mathrm{BCE}(\hat{y}^{\mathrm{img}}, y^{\mathrm{img}})$; during training, top-$k$ patches by $\hat{y}(P)$ may be sampled for stability ($k \in \mathbb{N}$, fixed across images).

(5) Stationarity regularization and spatial prior. For labeled normal patches, a stationarity/energy regularizer enforces contraction/non-accumulation:

$$\mathcal{L}_{\mathrm{stat}} = \mathbb{E}_{P\in\mathrm{normal}}\Big[\sum_{X\in\{R,M,Q\}} \big(E_X^{\mathrm{att}}(P) + \lambda_{\mathrm{ns}}\, \mathrm{NS}_X^{\mathrm{att}}(P)\big)\Big]. \tag{18}$$

A weak spatial prior promotes blockwise coherence in the pixel map $A(u)$ obtained from patches (defined as the average of scores over all patches covering u):

$$\mathcal{L}_{\mathrm{sp}} = \lambda_{\mathrm{tv}} \sum_{(u,v)\in\mathcal{N}} w_{uv}\, |A(u) - A(v)|, \tag{19}$$

where $\mathcal{N}$ is a fixed 4- or 8-neighborhood on Ω, and $w_{uv} \ge 0$ are edge weights derived from image/field gradients.

(6) Source objective.

$$\mathcal{L}_S = \mathcal{L}_{\mathrm{sup}} + \lambda_{\mathrm{stat}}\mathcal{L}_{\mathrm{stat}} + \lambda_{\mathrm{sp}}\mathcal{L}_{\mathrm{sp}}, \tag{20}$$

with $\mathcal{L}_{\text{sup}}$ the supervised loss on patch or image labels. The detector (g_ϕ, f_ψ), the time attention, and the spatial prior are *supervised* on $\mathcal{D}_S$. For feature calibration used later, estimate the mean μ_0 and covariance Σ_0 of $Z(P) = [E_R^{\text{att}}, E_M^{\text{att}}, E_Q^{\text{att}}, \text{NS}_R^{\text{att}}, \text{NS}_M^{\text{att}}, \text{NS}_Q^{\text{att}}, \text{DV}]^\top$ using *normal* source patches; denote by Σ_{Z_S} the covariance of Z over $\mathcal{D}_S$.

B. Target: unsupervised cross domain reuse.

(1) Target score and REF. Train a separate score $S_{\theta_T}(y,t) \approx \nabla_y \log p_t^{(T)}(y)$ on the unlabeled mixture $\mathcal{D}_T$ (same denoising objective). Build REF on $\mathcal{D}_T$ exactly as in the source, obtaining $\{\bar{R}_t(P), \bar{M}_t(P), \bar{Q}_t(P)\}$ and feature vectors: $Z_T(P) = [E_R^{\text{att}}, E_M^{\text{att}}, E_Q^{\text{att}}, \text{NS}_R^{\text{att}}, \text{NS}_M^{\text{att}}, \text{NS}_Q^{\text{att}}, \text{DV}]^\top$, computed with S_{θ_T}.

(2) Unsupervised CFA losses. CFA aligns target REF features to the *source field space* without target labels.

Time alignment learns a monotone reparameterization $\psi : \{1,\dots,T\} \to \{1,\dots,T\}$ minimizing

$$\mathcal{L}_{\text{time}} = \sum_{t=1}^{T} \left\| \mathbb{E}_{P\sim\mathcal{D}_T} \bar{R}_t - \mathbb{E}_{P\sim\mathcal{D}_S} \bar{R}_{\psi(t)} \right\|_2^2, \tag{21}$$

and applies $X_t \mapsto X_{\psi(t)}$ to all time-indexed target features. Here $\mathbb{E}_{P\sim\mathcal{D}_T}$ averages over patches from target domain.

Second order alignment whitens/recolors target features via $Z'_T = \Sigma_S^{1/2} \Sigma_T^{-1/2} (Z_T - m_T)$, where $m_T = \mathbb{E}Z_T$ and $\Sigma_T = \text{Cov}(Z_T)$ are the target mean/covariance, and $\Sigma_S = \text{Cov}(Z)$ is the source covariance of the same feature Z. A CORAL loss matches second moments and means:

$$\mathcal{L}_{\text{cor}} = \left\| \Sigma_{Z'_T} - \Sigma_{Z_S} \right\|_F^2 + \eta \left\| \mathbb{E}Z'_T - \mathbb{E}Z_S \right\|_2^2, \tag{22}$$

with $\|\cdot\|_F$ the Frobenius norm.

Directional alignment matches principal subspaces of stacked $\{\bar{M}_t\}_t$. Let $U_T, U_S \in \mathbb{R}^{d\times r}$ contain the top-r left singular vectors of the stacked sequences; an orthogonal Procrustes rotation $R \in \mathbb{R}^{r\times r}$ ($R^\top R = I$) minimizes

$$\mathcal{L}_{\text{sub}} = \|U_T R - U_S\|_F^2. \tag{23}$$

(3) Target objective and reuse.

$$\mathcal{L}_T = \mathcal{L}_{\text{time}} + \lambda_{\text{cor}} \mathcal{L}_{\text{cor}} + \lambda_{\text{sub}} \mathcal{L}_{\text{sub}}. \tag{24}$$

After optimizing CFA on $\mathcal{D}_T$, apply the learned ψ, the whitening/recoloring map, and the Procrustes rotation to target features, then run (g_ϕ, f_ψ) to produce patch probabilities $\hat{y}(P)$ and the pixel map $A(u)$. No target labels are required. *The risk bound supporting this reuse appears in Supp. Sec. S.7.*

C. End-to-end training protocol. *Stage S1* Train S_{θ_S} on $\mathcal{D}_S$ (mixture). *Stage S2* Build REF on $\mathcal{D}_S$, train (g_ϕ, f_ψ) with $\mathcal{L}_S$ (supervised), and estimate (μ_0, Σ_0) and Σ_{Z_S} using *normal* source patches (means/covariances of Z). *Stage T1* Train S_{θ_T} on $\mathcal{D}_T$ (mixture). *Stage T2* Build REF on $\mathcal{D}_T$, optimize $\mathcal{L}_T$ (unsupervised CFA). *Inference* Apply CFA to target features and run (g_ϕ, f_ψ) to obtain $A(u)$ and image-level scores (e.g., mean over top-$p\%$ pixels of A for a fixed $p \in (0, 100]$).

3.3. Part III: Summary and complexity

Summary. REF casts diffusion–based anomaly detection as a dynamical test of residual stationarity and persistence in a spatio–temporal vector field, allowing anomaly–related signal to be separated from stochastic and benign variation. Patch–time statistics over the field provide a practical anomaly score that respects the regional nature of real defects. CFA then aligns the field across domains in time, scale, and orientation, so that a detector calibrated on a labeled source can be reused on an unlabeled target.

Complexity. Let T be the number of diffusion steps, P the number of patches per image, D_f the Field Transformer token dimension, c_S the cost of one score evaluation, and K the average number of reverse–integration steps. For each image, REF extraction requires $O(P\,T\,c_S)$ to compute (R, M) and up to $O(P\,T\,K)$ for Q if full reverse trajectories are used; in practice, subsampling informative times keeps $K \ll T$. A lightweight L-layer transformer with standard self-attention incurs $O(P\,L\,(T^2 D_f + T D_f^2))$ time and $O(P\,T\,D_f)$ memory, which can be reduced to $O(P\,L\,T D_f^2)$ with linear–time attention variants. CFA operates on low-dimensional REF statistics Z (e.g., $m = \dim(Z) \approx 7$) and low-rank subspaces, so its cost is dominated by $O(nm^2)$ for covariance estimation with n patches and $O(m^3 + dr^2)$ for whitening/recoloring and Procrustes alignment (ambient dimension d, subspace rank r). Overall, the added REF + CFA overhead is modest compared to training the diffusion score networks, while remaining compatible with standard diffusion runtime budgets.

4. Experiments

4.1. Experiments Setup

Datasets and Source–Target Protocol. We evaluate on three benchmarks with fixed source–target pairings. For **MVTec** [4], the *bottle* category is used as the source domain, and *cable*, *capsule*, and *hazelnut* are used as target domains. In the **VisA** [59] setup, *candle* serves as the source domain, with *macaroni1*, *macaroni2*, and *pcb2* as target domains. For **DAGM** [41], *Class 2* is chosen as the source domain due to its strong in-domain performance, and *Class 1*, *Class 3*, and *Class 6* are used as target domains to construct more challenging transfer scenarios. In all cases, the source domain is fully labeled, whereas the target-domain training set is entirely unlabeled and anomaly-contaminated.

Table 1. AUROC (%) 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	76.70	82.50	81.40	77.80	66.00	76.40	68.30	74.30	95.30	77.63
	GLASS	74.80	76.50	77.60	94.94	78.37	84.31	80.19	79.48	70.66	79.65
	DDPM	69.71	69.34	71.85	83.95	76.21	73.08	80.11	75.77	77.43	75.27
	DDAD	73.91	75.60	76.47	91.69	89.10	77.04	86.00	87.81	82.27	82.21
Domain Adaptation	JWO	71.40	74.93	89.65	69.55	69.00	68.80	55.71	64.53	81.45	71.67
	PWAN	61.85	62.35	74.71	57.94	62.15	69.79	50.21	59.94	56.69	61.74
	MLWE	50.66	73.95	84.74	76.39	74.56	85.93	49.31	47.63	50.62	65.98
	CMKD	70.89	62.84	79.02	74.79	62.49	84.51	54.93	78.55	94.51	73.61
	UniNet	56.55	48.02	51.13	60.53	58.41	82.41	48.64	56.08	59.08	57.87
	ANC	54.78	59.17	40.05	57.53	57.50	53.47	56.89	57.63	51.96	54.33
	FFTAT	71.41	80.82	78.14	78.53	87.92	84.44	58.38	80.59	86.04	78.47
Domain Generalization	GGA	64.18	50.92	60.26	64.68	62.14	65.25	58.20	52.89	74.88	61.49
	BDC	48.91	50.00	53.46	44.17	48.19	42.50	51.53	56.81	62.30	50.87
	DDDG	65.74	63.40	80.34	70.12	65.50	79.38	57.65	56.73	63.53	66.93
	PMGDG	55.45	55.04	51.72	70.43	73.12	57.96	51.53	44.72	71.44	59.05
	DKGPL	72.63	67.57	87.58	74.12	67.45	56.45	50.15	60.63	58.72	66.14
Ours	REF+CFA	**81.72**	**85.13**	**91.66**	**100.00**	**99.50**	**98.95**	**100.00**	**100.00**	**100.00**	**95.22**
	Increase	*+5.02*	*+2.63*	*+2.01*	*+5.06*	*+10.40*	*+13.02*	*+14.00*	*+12.19*	*+4.70*	*+13.01*

Baseline. To ensure a thorough comparison with current state-of-the-art approaches, three baseline categories were evaluated. The unsupervised group comprised General-AD [33], GLASS [7], DDPM [18] and DDAD [26]. The domain adaptation group consisted of FFTAT [48], CMKD [58], UniNet [39], ANC [50], JWO [8], PWAN [37] and MLWE [40]. The domain generalisation group included BDC [52], DDDG [51], GGA [3], PMGDG [2] and DKGPL [38].

Evaluation Metrics. We assess the performance using the following three standard metrics: AUROC, AUPRC, and Rec@K. **AUROC** (area under the ROC) quantifies the model's ability to discriminate between normal and anomalous samples and serves as a stable, threshold-agnostic indicator of performance. **AUPRC** (area under the PRC) summarizes the precision-recall trade-off, which is particularly informative given the high class imbalance inherent in anomaly detection tasks. **Rec@K** (recall at rank K) measures the proportion of true anomalies retrieved among the top-K ranked instances, where K equals the number of anomalous samples in the test set. Owing to space constraints, AUROC is adopted as the primary metric and its results are reported in the main text.

More results about *AUPRC*, *Rec@K* are given in the *supplementary material*.

4.2. Comparison with SOTA

Table 1 reports AUROC scores on nine source–target transfers across MVTec, VisA, and DAGM. The proposed REF+CFA consistently achieves the best performance on all target domains, reaching an average AUROC of 95.22%, which exceeds the best average baseline (DDAD) by more than 13 percentage points (82.21%). On the three MVTec transfers (Bottle→Cable/Capsule/Hazelnut), REF+CFA improves over the best competing method by approximately 2–5 points, indicating a consistent gain even on relatively saturated benchmarks. On VisA, REF+CFA attains nearly perfect detection, achieving 100.00/99.50/98.95 on Macaroni1/Macaroni2/Pcb2, respectively, with gains of +5.06/+10.40/+13.02 points over the best competing methods. On DAGM, REF+CFA reaches 100.00% AUROC on Class 1, Class 3, and Class 6, respectively, while outperforming the best prior methods by +14.00/+12.19/+4.70 points. These results demonstrate that, across all three benchmark families, the proposed transfer-learning framework substantially outperforms both unsupervised anomaly detectors and state-of-the-art domain adaptation and domain generalization baselines.

4.3. Ablation Study

Table 2 reports ablations on three VisA transfers (macaroni1, macaroni2, pcb2), and Fig. 3 visualizes the cross-domain decoupling behaviour. The study examines the contributions of the REF, CFA, and REF-enabled decoupling of residual signals on target domains.

Effect of REF. Removing any REF component (R: gradient residual, M: directional offset, Q: path-integrated drift) degrades performance, showing that the three views are complementary. Dropping the directional term M causes the largest decline (e.g., -13.28 AUROC on *macaroni1*), indicating the importance of orientation informa-

Table 2. The results (AUROC, %) of the ablation study on different modules. The highest score is highlighted in **bold**.

Model	Macaroni1			Macaroni2			Pcb2			Average		
	AUROC	AUPRC	Rec@K	AUROC	AUPRC	Rec@K	AUROC	AUPRC	Rec@K	AUROC	AUPRC	Rec@K
w/o R	90.15	88.62	91.03	89.44	87.12	85.33	90.11	88.60	87.15	89.90	88.11	87.84
w/o M	86.72	85.11	87.20	85.79	84.95	83.93	87.46	86.72	85.14	86.66	85.59	85.42
w/o Q	87.03	86.40	88.15	86.13	85.02	84.08	88.03	87.01	86.30	87.06	86.14	86.18
w/o REF	84.11	82.30	85.00	83.19	81.95	82.50	85.02	83.12	84.00	84.11	82.46	83.83
w/o TA	96.03	95.14	96.33	95.10	94.01	89.02	96.55	93.88	90.89	95.89	94.34	92.08
w/o DSA	88.19	87.60	89.04	87.90	86.14	85.10	89.10	88.02	87.55	88.40	87.25	87.23
w/o SFA	92.44	90.81	93.10	91.60	90.11	91.02	93.15	91.04	91.89	92.40	90.65	92.00
w/o CFA	81.53	80.61	82.86	80.23	80.39	80.98	82.52	81.66	81.45	81.43	80.89	81.76
REF+CFA	**100.00**	**100.00**	**100.00**	**99.50**	**97.00**	**92.00**	**98.95**	**95.94**	**93.00**	**99.48**	**97.65**	**95.00**

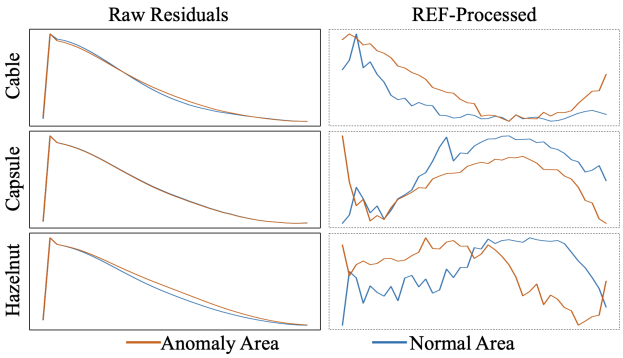


Figure 3. REF-enabled decoupling on target domains.

tion for separating anomaly-aligned signal from stochastic noise. Training directly on raw diffusion residuals (w/o REF) further reduces AUROC to 84.11/83.19/85.02, confirming that REF is crucial for stabilizing the anomaly signal under cross-domain transfer.

Effect of CFA. For cross-domain reuse, each CFA component makes a complementary contribution. Removing temporal alignment (w/o TA) yields moderate performance loss but remains significantly better than w/o REF, showing its importance in calibrating temporal drift. Directional subspace alignment (w/o DSA) is the most sensitive component to remove, resulting in large drops across datasets (e.g., AUROC 88.19 on *macaroni1*), consistent with the role of orientation alignment in preserving anomaly directionality. Second-order feature alignment (w/o SFA) also provides clear gains, especially on *pcb2*. When the full CFA module is disabled (w/o CFA), cross-domain transfer collapses (AUROC 80.23–82.52), demonstrating that alignment in temporal, directional, and second-order feature spaces is indispensable for reliable target-domain detection.

Residual Decoupling after Field Alignment. Figure 3 qualitatively illustrates REF's decoupling effect on target domains. Before processing, residual trajectories from normal and anomalous regions are closely overlapped and largely stochastic. After applying REF, the anomaly-aligned trend becomes more pronounced while normal regions contract toward stable evolution, yielding a clearer separation between the two.

4.4. Discussion and Future Work

The proposed framework demonstrates strong performance and robust generalization across diverse source–target pairs. By explicitly decomposing anomaly-aligned residual evolution (REF) and aligning spatio-temporal field representations across domains (CFA), the method maintains high accuracy even under substantial distribution shifts, enabling effective cross-domain reuse without target annotations. Despite these advantages, the primary limitation lies in computational cost. REF is built on the reverse dynamics of diffusion models, requiring score-based training in both the source and target domains. This increases training time and restricts scalability to settings where multiple domain pairs must be supported. A promising direction for future work is to reduce the time complexity of REF and CFA. Potential avenues include distilling the diffusion backbone into lightweight field predictors, exploring single-step residual evolution models, or designing amortized cross-domain alignment modules that operate without full diffusion training. Such developments could retain the benefits of field-based anomaly analysis while significantly lowering the computational burden.

5. Conclusion

This work presents a diffusion-residual framework for cross-domain label-free anomaly detection. The proposed REF captures persistent anomaly-related signals from residual evolution, while CFA aligns the resulting field space to enable transfer from labeled source domains to unlabeled target domains. In this way, anomaly-sensitive information can be separated from stochastic and intra-normal variability, enabling reliable cross-domain detection under substantial distribution shifts. Extensive experiments demonstrate that REF+CFA achieves state-of-the-art performance.

Acknowledgments

This work was supported by the Key R&D Program of Zhejiang (2025C02071) and the project Energy Data Real-Time Acquisition (Project No. 067).

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian conference on computer vision*, pages 622–637. Springer, 2018. 2

[2] Wen An, Zhixi Feng, and Shuyuan Yang. Progressive multiscale generator for domain generalization in hyperspectral image classification with small sample. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1, 2026. 7

[3] Aristotelis Ballas and Christos Diou. Gradient-guided annealing for domain generalization. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 20558–20568, 2025. 7

[4] 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 (CVPR)*, 2019. 6

[5] Zhangjie Cao, Lijia Ma, Mingsheng Long, and Jianmin Wang. Partial adversarial domain adaptation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 135–150, 2018. 3

[6] Zhangjie Cao, Kaichao You, Mingsheng Long, Jianmin Wang, and Qiang Yang. Learning to transfer examples for partial domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2985–2994, 2019. 3

[7] Qiyu Chen, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang. A unified anomaly synthesis strategy with gradient ascent for industrial anomaly detection and localization. In *European Conference on Computer Vision*, pages 37–54. Springer, 2024. 7

[8] Sentao Chen. Joint weight optimization for partial domain adaptation via kernel statistical distance estimation. *Neural Networks*, 180:106739, 2024. 7

[9] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020. 2

[10] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International conference on pattern recognition*, pages 475–489. Springer, 2021. 2

[11] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9737–9746, 2022. 2

[12] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7388–7398, 2022. 3

[13] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1705–1714, 2019. 2

[14] Mark S. Graham, Walter H. L. Pinaya, Petru-Daniel Tudosiu, Parashkev Nachev, Sebastien Ourselin, and M. Jorge Cardoso. Denoising diffusion models for out-of-distribution detection. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 2948–2957, 2023. 2

[15] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 98–107, 2022. 2

[16] 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 *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence*. AAAI Press, 2024. 2

[17] Dan Hendrycks, Mantas Mazeika, and Thomas Dietterich. Deep anomaly detection with outlier exposure. In *International Conference on Learning Representations*, 2019. 3

[18] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. 7

[19] 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. 2

[20] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9664–9674, 2021. 2

[21] 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. 2

[22] Yanghao Li, Naiyan Wang, Jianping Shi, Jiaying Liu, and Xiaodi Hou. Revisiting batch normalization for practical domain adaptation. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Workshop Track Proceedings*. OpenReview.net, 2017. 3

[23] Jian Liang, Dapeng Hu, Yunbo Wang, Ran He, and Jiashi Feng. Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(11):8602–8617, 2022. 3

[24] Zhenzhen Liu, Jin Peng Zhou, Yufan Wang, and Kilian Q Weinberger. Unsupervised out-of-distribution detection with diffusion inpainting. In *Proceedings of the 40th International Conference on Machine Learning*, pages 22528–22538. PMLR, 2023. 2

[25] 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. 3

[26] Arian Mousakhan, Thomas Brox, and Jawad Tayyub. Anomaly detection with conditioned denoising diffusion models. In *Pattern Recognition*, pages 181–195, Cham, 2025. Springer Nature Switzerland. 7

[27] Shuaicheng Niu, Jiaxiang Wu, Yifan Zhang, Yaofo Chen, Shijian Zheng, Peilin Zhao, and Mingkui Tan. Efficient test-time model adaptation without forgetting. In *International conference on machine learning*, pages 16888–16905. PMLR, 2022. 3

[28] Guansong Pang, Chunhua Shen, and Anton Van Den Hengel. Deep anomaly detection with deviation networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*, pages 353–362, 2019. 3

[29] 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. 2

[30] Lukas Ruff, Robert A. Vandermeulen, Nico Görnitz, Alexander Binder, Emmanuel Müller, Klaus-Robert Müller, and Marius Kloft. Deep semi-supervised anomaly detection. In *International Conference on Learning Representations*, 2020. 3

[31] Kuniaki Saito, Shohei Yamamoto, Yoshitaka Ushiku, and Tatsuya Harada. Open set domain adaptation by backpropagation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 153–168, 2018. 3

[32] 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. 2

[33] Luc PJ Sträter, Mohammadreza Salehi, Efstratios Gavves, Cees GM Snoek, and Yuki M Asano. Generalad: Anomaly detection across domains by attending to distorted features. In *European Conference on Computer Vision*, pages 448–465. Springer, 2024. 7

[34] Yu Sun, Xiaolong Wang, Zhuang Liu, John Miller, Alexei Efros, and Moritz Hardt. Test-time training with self-supervision for generalization under distribution shifts. In *International conference on machine learning*, pages 9229–9248. PMLR, 2020. 3

[35] Dequan Wang, Evan Shelhamer, Shaoteng Liu, Bruno Olshausen, and Trevor Darrell. Tent: Fully test-time adaptation by entropy minimization. In *International Conference on Learning Representations*, 2021. 3

[36] Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for anomaly detection. *Proceedings of the British Machine Vision Conference 2021*, 2021. 2

[37] Zi-Ming Wang, Nan Xue, Ling Lei, Rebecka Jörnsten, and Gui-Song Xia. Partial distribution matching via partial wasserstein adversarial networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(9):7944–7959, 2025. 7

[38] Jiacheng Wei, Qibin Wang, Guowei Zhang, Hongbo Ma, and Yi Wang. Domain knowledge guided pseudo-label generation framework for semi-supervised domain generalization fault diagnosis. *Advanced Engineering Informatics*, 67: 103540, 2025. 7

[39] Shun Wei, Jielin Jiang, and Xiaolong Xu. Uninet: A contrastive learning-guided unified framework with feature selection for anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 9994–10003, 2025. 7

[40] Lisheng Wen, Sentao Chen, Zijie Hong, and Lin Zheng. Maximum likelihood weight estimation for partial domain adaptation. *Information Sciences*, 676:120800, 2024. 7

[41] Matthias Wieler, Tobias Hahn, and Fred A. Hamprecht. Weakly supervised learning for industrial optical inspection, 2007. Dataset. 6

[42] Julian Wyatt, Adam Leach, Sebastian M. Schmon, and Chris G. Willcocks. Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 650–656, 2022. 2

[43] Haohao Xu, Shuchang Xu, and Wenzhen Yang. Unsupervised industrial anomaly detection with diffusion models. *Journal of Visual Communication and Image Representation*, 97:103983, 2023. 2

[44] Xincheng Yao, Ruoqi Li, Jing Zhang, Jun Sun, and Chongyang Zhang. Explicit boundary guided semi-push-pull contrastive learning for supervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24490–24499, 2023. 3

[45] Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian conference on computer vision*, 2020. 2

[46] Kaichao You, Mingsheng Long, Zhangjie Cao, Jianmin Wang, and Michael I Jordan. Universal domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2720–2729, 2019. 3

[47] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677*, 2021. 2

[48] Xiaowei Yu, Zhe Huang, and Zao Zhang. Feature fusion transferability aware transformer for unsupervised domain

adaptation. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 6752–6761, 2025. 7

[49] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draema discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 8330–8339, 2021. 2

[50] Chi Zhang, Cheng Jingpu, Yanyu Xu, and Qianxiao Li. Parameter-efficient fine-tuning with controls. In *Forty-first International Conference on Machine Learning*, 2024. 7

[51] Guowei Zhang, Xianguang Kong, Hongbo Ma, Qibin Wang, Jingli Du, and Jinrui Wang. Dual disentanglement domain generalization method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, 228:112460, 2025. 7

[52] Guowei Zhang, Xianguang Kong, Qibin Wang, Jingli Du, and Hongbo Ma. Balancing discrepancy and consistency: Adversarial single domain generalization in fault diagnosis. *IEEE Transactions on Industrial Informatics*, 21(10):7923–7934, 2025. 7

[53] Jing Zhang, Zewei Ding, Wanqing Li, and Philip Ogunbona. Importance weighted adversarial nets for partial domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8156–8164, 2018. 3

[54] Xinyi Zhang, Naiqi Li, Jiawei Li, Tao Dai, Yong Jiang, and Shu-Tao Xia. Unsupervised surface anomaly detection with diffusion probabilistic model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6782–6791, 2023. 2

[55] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3914–3923, 2023. 2

[56] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16699–16708, 2024. 3

[57] Qihang Zhou, Guansong Pang, Yu Tian, Shibo He, and Jiming Chen. AnomalyCLIP: Object-agnostic prompt learning for zero-shot anomaly detection. In *The Twelfth International Conference on Learning Representations*, 2024. 2

[58] Wenlve Zhou and Zhiheng Zhou. Unsupervised domain adaption harnessing vision-language pre-training. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(9):8201–8214, 2024. 7

[59] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pretraining for anomaly detection and segmentation. In *European conference on computer vision*, pages 392–408. Springer, 2022. 6