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.

GPFlow: Gaussian Prototype Probability Flow for Unsupervised Multi-Modal Anomaly Detection

Yiting Li[1], Xulei Yang[1]*, Jingyi Liao[1], Jing Zhang[2], Fayao Liu[1]*
[1]Institute for Infocomm Research, A*STAR, Singapore
[2]School of Computer Science and Engineering, Macao University of Science and Technology
{li_yiting, yang_xulei, liao_jingyi, liu_fayao}@i2r.a-star.edu.sg

Abstract

In this paper, we study unsupervised multi-modal anomaly detection under challenging few-shot conditions, where only a few normal training samples are available for each class. To prevent the trivial reconstruction of anomalies, recent methods often rely on discrete prototypes to establish an information bottleneck. However, such discrete prototypes fail to capture continuous and anisotropic variations of normal features. We therefore propose GPFlow, a probability-flow-inspired framework that models normality using learnable Gaussian prototypes. The key component of GPFlow is an analytical Posterior-Mean Path (PMP) router, which reconstructs features through the posterior mean of a noise-smoothed Gaussian mixture. This yields anisotropic shrinkage as a covariance-aware information bottleneck: PMP selectively preserves normal variations aligned with the covariance structure of Gaussian prototypes while strictly suppressing deviations inconsistent with the prototypes. To exploit complementary knowledge across modalities, GPFlow further combines intra-modal and cross-modal reconstruction, and applies a lightweight instance-aware prior calibration to alleviate the distribution mismatch between sparse training data and diverse test samples. Experiments on MVTec-3D-AD and Eyecandies show that GPFlow achieves significant performance improvement with only a few normal training samples while remaining computationally efficient.

1. Introduction

Multimodal anomaly detection (MAD) [6, 19, 41] aims at identifying defects by jointly inspecting multiple modalities, such as RGB images, 3D point clouds, and surface normal maps, which is becoming increasingly important for industrial quality control [17, 45]. By combining visual appearance with shape information, MAD can capture subtle defects that would remain invisible to a single modality, thereby providing a more comprehensive inspection of products. However, achieving robust MAD in unsupervised few-shot normal settings remains highly challenging. In many real scenarios, only a handful of defect-free samples per class are available for model training due to the high cost of acquiring such multimodal data, making reliable anomaly detection particularly challenging.

Existing reconstruction-based MAD approaches leverage multiple modalities to reconstruct normal features and localize defects from reconstruction discrepancies. For example, 3D-ADNAS [22] searches for effective multimodal fusion architectures, while FIND [19] combines reverse distillation with cross-modal feature transformation. However, these methods often lack a sufficiently strong information bottleneck and thus suffer from the identity shortcut issue [7, 8, 16, 44], where a powerful network can simply reproduce its inputs (including anomalous patterns) with minimal reconstruction error. To address this issue, recent methods [20, 24, 26] explore discrete prototypes to enforce a stronger information bottleneck. Although such discrete routing restricts the reconstruction of anomalies, it still relies on hard quantization and isotropic distance metrics for prototype assignment. Consequently, this fails to capture the continuous and anisotropic variations of normal distribution, causing normal features that vary along high-variance directions to incur large reconstruction errors, while anomalous features along low-variance dimensions remain insufficiently penalized. We illustrate this issue with a toy example in Fig. 2.

The above limitation leads to a key question: *how can we compactly capture the continuous distribution of normal appearance and geometry, while enforcing a covariance-aware information bottleneck to robustly suppress anomalies?* Towards this goal, we introduce GPFlow, a multi-modal reconstruction framework that represents normality using Gaussian prototypes [21, 34, 40, 43], where each prototype is parameterized by a learnable mean, diagonal covariance, and mixture weight. This allows GPFlow

*Corresponding authors.

to utilize a distribution-level description of normality rather than patch-level biased normality information. The core idea of GPFlow is a probability flow inspired by Posterior Mean Path (PMP) router [12, 23, 35]: PMP applies a sequence of deterministic updates on input features toward the posterior mean of a noise-smoothed Gaussian mixture [15, 33]. Intuitively, this moves input features into the high-probability neighborhoods of the prototypes (the normal manifold), while suppressing out-of-distribution variations [9, 37, 39]. Crucially, the PMP router leverages the covariance structure of the Gaussian prototypes to perform anisotropic shrinkage; it selectively preserves variations aligned with the learned covariance structure while suppressing deviations in other directions. The routing thus acts as a covariance-aware information bottleneck to prevent trivial reconstruction of anomalies: features consistent with normal variability will remain stable during updating, whereas features inconsistent with the prototypes are strongly pulled toward their nearest normal modes. The resulting residuals between inputs and prototype-constrained reconstructions thus produce strong reconstruction errors on anomalous regions and enable accurate anomaly detection.

To fully exploit complementary appearance and geometric information, GPFlow integrates a PMP-constrained coupled reconstruction design that routes input features through both intra-modal and cross-modal pathways. Furthermore, since models trained on few-shot data inevitably face distribution shifts between training and testing samples [4, 25, 28, 42], we perform an instance-aware prior calibration during inference. This step dynamically adjusts the Gaussian prototypes using confidently normal patches selected from the current test sample, allowing the learned priors to cover unseen normal variations without being contaminated by anomalous features. To summarize, our main contributions are:

- We propose GPFlow, a probability-flow-inspired multimodal anomaly detection framework that models continuous normal variations with Gaussian prototypes and reconstructs inputs via an analytical Posterior-Mean Path (PMP).
- We introduce anisotropic shrinkage as a covariance-aware information bottleneck, which effectively prevents the trivial reconstruction of anomalies.
- Extensive experiments on MVTec-3D-AD and Eyecandies datasets demonstrate that GPFlow achieves state-of-the-art anomaly detection performance under challenging few-shot settings.

2. Related Work

Multimodal Anomaly Detection. Recent industrial anomaly detection research has extended from conventional RGB+3D settings to multi-sensor scenarios, such as MulSen-AD [17] and Real-IAD D3 [45]. Existing MAD strategies can be summarized into memory-based retrieval, cross-modal alignment, and multimodal reconstruction paradigms. Memory-based approaches such as M3DM [41] and SG-DM [5] maintain a repository of normal feature exemplars and detect anomalies by measuring deviations from this memory. Cross-modal alignment methods such as CFM [6] and LSFA [38] aim to learn explicit correspondences between modalities. Reconstruction-based methods leverage multiple modalities for joint reconstruction. 3D-ADNAS [22] optimizes fusion architectures via neural architecture search, while FIND [19] combines reverse distillation with cross-modal feature mapping.

Prototype-Based Anomaly Detection. Recent studies have also explored prototype-based designs for 2D anomaly detection. HVQ-Trans [24] learns a hierarchical vector-quantized codebook of representative normal patterns and reconstructs inputs through discrete prototype assignments. INP-Former [26] derives Intrinsic Normal Prototypes directly from the normal regions of each test image and uses them as on-the-fly references for reconstruction. PIRN [20] introduces a prototype-driven reconstruction framework with balanced optimal transport for prototype assignment. Despite their strong performance, both HVQ-Trans and INP-Former rely on discrete prototypes, which limits their ability to capture continuous variations of normal patterns and lacks a Bayesian MMSE interpretation. By contrast, our PMP router is a covariance-aware information bottleneck, preserving normal variability while suppressing out-of-distribution deviations. In the supervised open-set setting, DPDL [40] constructs learnable Gaussian prototypes and trains a Schrödinger bridge diffusion process to align normal samples with the prototypes while separating anomalies. While effective, DPDL requires access to a small number of real anomalous samples for training. In contrast, our framework only requires a few normal samples.

3. Method

3.1. Framework Overview

Fig. 1 shows an overview of GPFlow, which takes RGB images I_{rgb} as *appearance* modality and corresponding 3D inputs (i.e., surface normal maps) I_{shape} as *shape* modality. Given the scarcity of normal training data, we utilize pre-trained encoders (e.g., DINOv2) to extract multi-scale features for both modalities. These features are aggregated into patch tokens $T_{\text{rgb}}, T_{shape} \in \mathbb{R}^{B\times N\times D}$, where B, N, D denote the batch size, number of patches, and the feature dimension.

Gaussian Prototype Modeling. To capture the continuous variation of normality under few-shot constraints, GPFlow models the feature distribution of each modality as a Gaussian Mixture Model (GMM) parameterized by K learnable

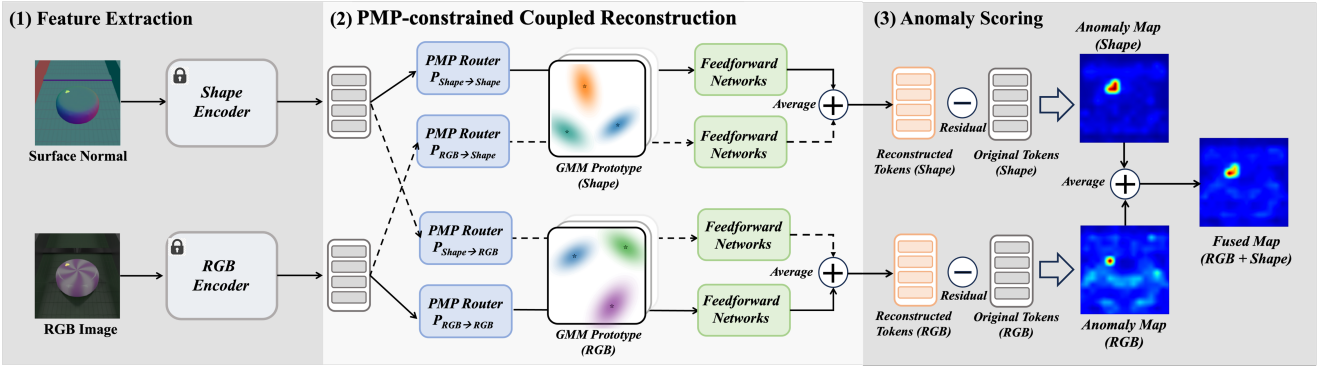

Figure 1. **Overview of the proposed GPFlow framework.** GPFlow operates in three main stages. (1) **Feature Extraction:** Frozen pre-trained encoders extract features from RGB (appearance) and Surface Normal (shape) inputs. (2) **PMP-constrained Coupled Reconstruction:** This module models the distribution of normality using learnable Gaussian Mixture Model (GMM) prototypes. The Posterior-Mean Path (PMP) router iteratively aligns input features with these prototypes via both intra-modal (solid lines) and cross-modal (dashed lines) pathways, enforcing comprehensive multi-modal consistency. (3) **Anomaly Scoring:** Anomaly maps are generated by computing the discrepancy between the original input tokens and the prototype-constrained reconstructions, where the PMP router serves as a covariance-aware information bottleneck to suppress the reconstruction of anomalies.

prototypes: $\mathcal{P} = \{(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k, \pi_k)\}_{k=1}^K$ (mean $\boldsymbol{\mu}$, covariance $\boldsymbol{\Sigma}$, and mixture weight π). We assume diagonal covariance matrices $\boldsymbol{\Sigma}_k$ for computational efficiency. These prototypes define a GMM prior $p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, providing a compact and continuous description of normality. Building upon this probabilistic foundation, GPFlow comprises the following three key components:

3.2. The Posterior-Mean Path (PMP) Router

The core challenge in reconstruction-based MAD models is enforcing an information bottleneck to prevent the "identity shortcut" issue, i.e., trivial reconstruction of anomalous inputs. We address this issue by introducing the Posterior-Mean Path (PMP) router, a mechanism inspired by probability flow and score matching [37, 39], which iteratively guides features toward high-probability regions of the learned Gaussian prototype distribution.

Score Ascent and Probability-Flow View. Consider an observation $y = x+\epsilon$, where x is a clean feature drawn from the normal distribution $p(x)$, and $\epsilon \sim \mathcal{N}(0, \tau^2\mathbf{I})$ is Gaussian noise. The Minimum Mean Squared Error (MMSE) estimate of x is the posterior mean $\mathbb{E}[x|y]$ [11, 13]. Tweedie's formula [13, 27, 29] establishes a crucial link between the posterior mean and the score:

$$\mathbb{E}[x|y] = y + \tau^2 \nabla_y \log p_\tau(y), \tag{1}$$

where $p_\tau(y)$ is the noise-smoothed density. Crucially, the score function $\nabla_y \log p_\tau(y)$ points toward regions of higher probability density [36, 37]. Hence, moving a sample y toward its posterior mean is equivalent to taking a score-ascent step on the smoothed log-density $\log p_\tau(y)$ [9, 39].

At a fixed noise level τ, the continuous-time limit of these score-ascent steps yields the deterministic flow

$$\dot{x}(t) = \tau^2 \nabla_x \log p_\tau(x(t)), \tag{2}$$

which can be viewed as a probability-flow-inspired trajectory toward high-density regions of the learned normal distribution [23, 37]. In the AD task, this provides a fixed-noise score-ascent path for denoising anomalous features toward the learned normal manifold.

Closed-Form Posterior Mean with GMM Priors. While simulating a true continuous flow typically requires solving ODEs, the choice of a GMM prior yields an efficient analytical solution. Let $p(x) = \sum_{k=1}^K \pi_k \mathcal{N}(x; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$. Then the smoothed density is also a GMM, $p_\tau(y) = \sum_{k=1}^K \pi_k \mathcal{N}(y; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k + \tau^2\mathbf{I})$, and the posterior mean admits a closed form [3, 27]:

$$D_\tau(y) \triangleq \mathbb{E}[x|y] = \sum_{k=1}^K \gamma_k(y) m_k(y), \tag{3}$$

which is a weighted average of component-wise posterior means. This contains two distinct parts: soft assignment across prototypes and shrinkage within each prototype.

The weighting factor $\gamma_k(y)$ is the *responsibility* (soft assignment) of component k for observation y, determining *which* prototypes contribute to the reconstruction:

$$\gamma_k(y) = \frac{\pi_k \mathcal{N}(y; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k + \tau^2\mathbf{I})}{\sum_{j=1}^K \pi_j \mathcal{N}(y; \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j + \tau^2\mathbf{I})}. \tag{4}$$

The term $m_k(y)$ is the *component posterior mean*, i.e., the optimal MMSE estimate assuming that y is generated from the k-th prototype [31]:

$$m_k(y) = \mathbf{A}_k y + (\mathbf{I} - \mathbf{A}_k)\boldsymbol{\mu}_k, \qquad \mathbf{A}_k = \boldsymbol{\Sigma}_k(\boldsymbol{\Sigma}_k + \tau^2\mathbf{I})^{-1}. \tag{5}$$

The corresponding component posterior covariance is $\mathbf{C}_k = (\mathbf{\Sigma}_k^{-1} + \tau^{-2}\mathbf{I})^{-1} = \tau^2\mathbf{A}_k$.

From Component-Wise Shrinkage to the Mixture Posterior Mean. Eq. (5) describes the shrinkage behavior within each Gaussian component. However, PMP in Eq. (3) is an input-dependent mixture over components through the responsibilities $\gamma_k(y)$. Therefore, the local geometry of PMP is the Jacobian of the mixture posterior-mean $D_\tau(y)$.

Using Tweedie's formula together with the posterior covariance identity [27], we obtain

$$\mathbf{J}_{D_\tau}(y) = \mathbf{I} + \tau^2\nabla_y^2 \log p_\tau(y) = \tau^{-2}\,\mathrm{Cov}[x|y]. \tag{6}$$

For the GMM posterior, the law of total covariance gives

$$\mathrm{Cov}[x|y] = \sum_{k=1}^{K} \gamma_k(y)\Big[\mathbf{C}_k + \Delta_k(y)\Delta_k(y)^\top\Big]. \tag{7}$$

Here, $\Delta_k(y) \triangleq m_k(y) - D_\tau(y)$ denotes the deviation of the k-th component posterior mean from the full posterior mean. Substituting $\mathbf{C}_k = \tau^2\mathbf{A}_k$ into Eq. (6) yields

$$\mathbf{J}_{D_\tau}(y) = \sum_{k=1}^{K} \gamma_k(y)\mathbf{A}_k + \tau^{-2}\sum_{k=1}^{K} \gamma_k(y)\Delta_k(y)\Delta_k(y)^\top. \tag{8}$$

The term $\sum_k \gamma_k(y)\Delta_k(y)\Delta_k(y)^\top$ captures the between-prototype dispersion induced by the disagreement among multiple prototypes. Eq. (8) reveals two complementary effects. The first term is the *within-prototype anisotropic shrinkage* encoded by $\mathbf{A}_k$, while the second term is a *between-prototype ambiguity term* induced by competition among multiple prototypes. Consequently, PMP behaves like a locally contractive shrinkage when one prototype dominates the posterior. In contrast, near the transition region between prototypes, the between-prototype ambiguity term becomes stronger, so the update path is jointly determined by multiple components.

Anisotropic Shrinkage as Information Bottleneck. In particular, when a dominant prototype k explains the observation, $\mathbf{A}_k$ acts as a dimension-wise "retention coefficient," determining the convex combination between the observation y and the prototype mean $\boldsymbol{\mu}_k$. Since $\mathbf{\Sigma}_k = \mathrm{diag}(\sigma_{k,c}^2)$ is diagonal, $\mathbf{A}_k$ is also diagonal with elements

$$a_{k,c} = \frac{\sigma_{k,c}^2}{\sigma_{k,c}^2 + \tau^2} \in (0, 1).$$

This achieves *anisotropic shrinkage* [27]: it preserves feature variations along dimensions where the prototype exhibits high variance (large $\sigma_{k,c}^2$, gain $\to 1$), while strongly suppressing anomalous variations(small $\sigma_{k,c}^2$, gain $\to 0$). As such, the gain matrix $\mathbf{A}_k$ controls the strength of shrinkage along different feature dimensions within the k-th prototype, acting as a covariance-aware information bottleneck.

To intuitively see why this forms an information bottleneck, we can rewrite Eq. (5) to focus on the displacement:

$$y - m_k(y) = (\mathbf{I} - \mathbf{A}_k)(y - \boldsymbol{\mu}_k). \tag{9}$$

This shows that shrinkage is proportional to the deviation from the prototype mean, but modulated dimension-wise by the prototype covariance. In this way, PMP performs a soft projection that preserves variations aligned with the learned covariance structure while suppressing deviations in unaligned directions.

Iterative PMP Routing as Score Ascent. The PMP router iteratively applies this analytical posterior mean update. Instead of directly solving a continuous ODE, it defines a discrete trajectory toward the normal manifold modeled by the Gaussian prototypes:

$$x^{(t+1)} = (1 - \beta_t)x^{(t)} + \beta_t\mathbb{E}[x|y = x^{(t)}]. \tag{10}$$

Substituting Eq. (1) into Eq. (10) gives

$$x^{(t+1)} = x^{(t)} + \eta_t\nabla_x \log p_\tau(x^{(t)}), \qquad \eta_t = \beta_t\tau^2. \tag{11}$$

Therefore, PMP is a fixed-noise score-ascent step on the smoothed density p_τ. We further quantify in the appendix how the routing parameters (τ, β_t, T) determine the strength of PMP.

Comparison with Discrete Prototype Methods. The key difference between PMP and discrete prototype methods [20, 24, 26] lies in how an input feature is reconstructed. A discrete router replaces each token with a single nearest prototype or a convex combination of prototypes in a discrete codebook. Therefore, its routing displacement is determined by the Euclidean distance to the selected prototype centers. This can be problematic when the actual normal feature distribution is anisotropic: normal features that vary along high-variance directions can still incur large reconstruction errors, even when they remain in high-density regions of the normal distribution. In contrast, a deviation along a low-variance direction is more likely to be anomalous, but may not be penalized sufficiently due to smaller routing displacement in Euclidean distance.

By contrast, PMP avoids this limitation by using the GMM posterior mean in Eq. (3). As shown by Eqs. (5) and (9), the routing displacement is modulated by $(\mathbf{I} - \mathbf{A}_k)$, which produces covariance-aware anisotropic shrinkage: Normal variations along high-variance directions are largely preserved. In contrast, anomalous deviations along low-variance directions are strongly suppressed. Moreover, from Eq. (1) and the iterative update in Eq. (10), PMP is a

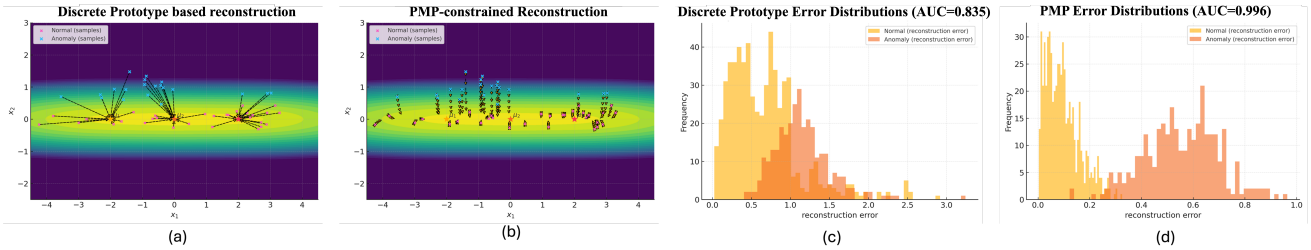

Figure 2. Conceptual comparison between Discrete Prototype methods and our Posterior-Mean Path (PMP) routing in a 2D three-component Gaussian mixture. Normal samples are drawn from a three-component Gaussian mixture with means $\boldsymbol{\mu}_1 = (-2, 0)$, $\boldsymbol{\mu}_2 = (0, 0)$, $\boldsymbol{\mu}_3 = (2, 0)$ and diagonal covariances $\boldsymbol{\Sigma}_1 = \mathrm{diag}(1.2, 0.03)$, $\boldsymbol{\Sigma}_2 = \mathrm{diag}(1.0, 0.03)$, $\boldsymbol{\Sigma}_3 = \mathrm{diag}(1.2, 0.03)$. Anomalies are generated by vertically shifting normal samples. PMP is run with $T = 5$ steps and observation noise variance $\tau^2 = 0.05$. **(a)** Discrete Prototype methods incur large routing displacements for normal samples varying along the high-variance axes. **(b)** PMP keeps normal samples geometrically stable while strongly pulling anomalies toward high-density normal modes. **(c-d)** Resulting reconstruction error distributions. PMP achieves a significantly clearer separation between normal and anomalous samples, improving the anomaly detection AUC from 0.835 to 0.996.

discrete score-ascent process on the smoothed density. Consequently, normal samples already near high-density normal modes undergo only small routing displacement, while anomalous samples in low-density regions are pulled more significantly. This leads to reconstruction errors that are better aligned with the learned normal distribution. We further illustrate this behavior with a toy example in Fig. 2: discrete prototype methods can over-penalize normal samples that vary along an anisotropic high-variance direction, whereas PMP preserves such normal variation and yields a larger separation between normal and anomalous samples.

3.3. Instance-Aware Prior Calibration (IPC)

Sparse training data may not fully cover the diversity of normal test samples, causing unseen-yet-normal patterns to be over-penalized by a fixed Gaussian prior. Inspired by [26], we perform a lightweight *instance-aware prior calibration* before PMP routing. Importantly, this calibration is conducted independently for each test sample and does not accumulate across the test set.

Confidence-Aware Patch Filtering. We first identify confidently normal patches within the input features (e.g., T_{rgb}) using the Mutual Scoring Mechanism (MSM) [18]. A binary mask M is formed by selecting the lowest-scoring ρ fraction of patches, such that only reliable normal tokens participate in prior calibration. The posterior responsibilities in Eq. (4) are then masked as $\gamma^{\mathrm{M}} = \gamma \odot M$. For each prototype k, we compute the responsibility-weighted prototype evidence

$$U_k = \frac{\sum_{n=1}^{N} \gamma_{nk}^{\mathrm{M}} t_n}{\sum_{n=1}^{N} \gamma_{nk}^{\mathrm{M}}}, \tag{12}$$

where t_n denotes the n-th patch token and γ_{nk}^{M} is the masked responsibility of prototype k for t_n.

Prior Calibration. Given the prototype evidence U_k, we calibrate the prototype mean using an exponential moving average (EMA): $\boldsymbol{\mu}_k' = (1 - \eta_\mu)\boldsymbol{\mu}_k + \eta_\mu U_k$,

where $\eta_\mu \in (0, 1)$ controls the strength of instance-specific recentering. We then estimate the responsibility-weighted sample variance around the calibrated mean:

$$\hat{\mathbf{v}}_k = \frac{\sum_{n=1}^{N} \gamma_{nk}^{\mathrm{M}} (t_n - \boldsymbol{\mu}_k')^2}{\sum_{n=1}^{N} \gamma_{nk}^{\mathrm{M}}}. \tag{13}$$

To ensure positivity and numerical stability, the diagonal covariance is updated in the log-variance domain:

$$\log(\boldsymbol{\Sigma}_k') = (1 - \eta_\Sigma) \log(\boldsymbol{\Sigma}_k) + \eta_\Sigma \log(\hat{\mathbf{v}}_k + \epsilon), \tag{14}$$

where $\eta_\Sigma \in (0, 1)$ is the covariance momentum and $\epsilon > 0$ is a small constant. The calibrated Gaussian prior $(\boldsymbol{\mu}_k', \boldsymbol{\Sigma}_k')$ is used only for the current sample in the subsequent PMP routing, and is discarded afterward.

3.4. PMP-constrained Coupled Reconstruction (PCR)

Coupled reconstruction. Following the coupled reconstruction paradigm in prior works, GPFlow combines an intra-modal pathway and a cross-modal pathway under the same PMP router. This design enforces both modality-specific consistency and cross-modal coherence by utilizing the PMP router. We define the PMP routing operation as $R(T; \mathcal{P})$, which takes input tokens T and routes them toward the manifold defined by prototypes $\mathcal{P}$ using the iterative updates in Eq. (10).

Using the RGB branch as an example, the intra-modal and cross-modal reconstructions for the RGB branch are defined as:

$$\hat{T}_{\mathrm{rgb}}^{\mathrm{intra}} = R(T_{\mathrm{rgb}}; \mathcal{P}_{\mathrm{rgb}}), \qquad \hat{T}_{\mathrm{shape}\to\mathrm{rgb}}^{\mathrm{cross}} = R(T_{\mathrm{shape}}; \mathcal{P}_{\mathrm{rgb}}). \tag{15}$$

They are further aggregated by a lightweight FFN:

$$T'_{\text{rgb}} = \text{FFN}_{\text{rgb}}\big(\hat{T}^{\text{intra}}_{\text{rgb}}\big) + \text{FFN}_{\text{rgb}}\big(\hat{T}^{\text{cross}}_{\text{shape}\to\text{rgb}}\big). \quad (16)$$

The shape branch T_{shape} is defined in the same way. With the anisotropic shrinkage bottleneck, the decoder learns to reconstruct normal patterns consistent with the Gaussian prototypes, ensuring that anomalous patterns result in larger reconstruction errors.

Anisotropic Shrinkage Enlarges Anomalous Displacement. A key difference from conventional coupled reconstruction methods such as FIND [19] is that GPFlow explicitly enlarges the displacement gap between normal and anomalous inputs before decoding. From Eq. (5) and Eq. (9), when an input token can be explained by one Gaussian prototype, the routing displacement is controlled by $(\mathbf{I} - \mathbf{A}_k)$. Therefore, PMP applies covariance-aware anisotropic shrinkage: normal variations along high-variance directions are largely preserved, while anomalous deviations produce larger routing displacements in both the intra-modal and cross-modal branches. This explicitly makes anomalous patterns harder to reconstruct.

Training Objective. GPFlow is optimized by enforcing a cosine similarity based soft mining loss [26] between the raw features extracted by the encoders and the coupled reconstructions. At inference, anomaly maps are computed from the cosine similarity based discrepancies between the encoder features and the final coupled reconstructions. The modality-specific anomaly maps are then fused to produce the final anomaly prediction.

4. Experiment

Datasets. We evaluate GPFlow on the MVTec-3D-AD [1] and Eyecandies [2] datasets under data-scarce conditions by randomly sampling 5, 10 or 50 normal images per class. For each k-shot setting, we repeat the random selection 10 times and report the average performance to mitigate selection bias.

Implementation Details. We employ two frozen DINOv2 ViT-B/14 [30] as backbone encoders for the RGB and shape modalities, with inputs resized to 224×224. Patch tokens are from layers 2 through 9, and we discard the class tokens and average the selected layers to form per-patch tokens. We utilize $K = 30$ Gaussian prototypes per modality. For GMM initialization, means $\boldsymbol{\mu}$ are initialized from $\mathcal{N}(0, 1)$ and log-variances $\log \Sigma$ are initialized to -2.0. During reconstruction, the PMP router operates for 8 steps with a fixed observation noise $\tau^2 = 10^{-2}$ and a linear step schedule $\beta \in [0.2, 0.8]$. For the instance-aware prior calibration (IPC), we use a momentum of 0.1 for the EMA update. For each image, we keep only the most consistent 60% patches ($\rho = 0.6$). We train the model per category using the Adam optimizer [14] with a learning rate of 1×10^{-4}. Anomaly maps are generated by summing the RGB and shape reconstruction errors, followed by smoothing with a 5×5 Gaussian kernel. Image-level scores are the maximum value of the merged anomaly map.

4.1. Main Results

We compare GPFlow against several leading unsupervised MAD methods, including memory-based approaches (M3DM [41]), alignment-based methods (CFM [6]), reconstruction-based methods (AST [32], BTF [10], 3D-ADNAS [22]), and the most recent approach FIND [19], with results summarized in Tab. 1.

Few-Shot Anomaly Detection. GPFlow demonstrates significant superiority under few-shot settings. In the extreme 5-shot setting on MVTec-3D-AD, GPFlow achieves an I-AUROC of 0.912, outperforming the second-best method, FIND (0.899), and significantly surpassing established approaches like CFM (0.811) and M3DM (0.822). The advantage is even more pronounced on the challenging Eyecandies dataset. In the 5-shot setting, GPFlow achieves an I-AUROC of 0.913, establishing a substantial margin over FIND (0.840) and CFM (0.795). As the number of training shots increases to 10 and 50, GPFlow consistently maintains its lead in I-AUROC on both datasets. For instance, in the 10-shot MVTec-3D-AD setting, GPFlow achieves 0.935 I-AUROC. This robust performance under limited data validates the effectiveness of GPFlow in modeling normality with Gaussian prototypes.

Full-Data Anomaly Detection. When trained with all available normal data (All-shot setting) on MVTec-3D-AD, GPFlow achieves a competitive I-AUROC of 0.970, slightly behind FIND (0.978). This illustrates a trade-off: while GPFlow's strong inductive biases (Gaussian prototypes and PMP router) ensure superior few-shot robustness by imposing an explicit information bottleneck, FIND's multi-stage architecture works better with abundant data, allowing it to capture finer-grained variations than our computationally efficient framework.

Visualization of PMP as an Information Bottleneck. To validate the effectiveness of the PMP router as an information bottleneck, we visualize the feature displacements induced during the routing process. We illustrate the log density of the learned Gaussian prototypes in a 2D space. As shown in Tab. 1(b), normal features (green) align well with the Gaussian prototype manifold and lie in high-density regions. Therefore, they undergo minimal displacement during the PMP routing process. In contrast, anomalous features (red) are pulled along the probability gradient toward the nearest high-density prototypes.

4.2. Ablation Studies

We conduct extensive ablation studies to analyze the contributions of each component, mainly under the 10-shot set-

Table 1. (a) Comparison of anomaly detection and localization performance on MVTec-3D-AD and Eyecandies with varying numbers of training shots. **Bold** indicates the best performance. (b) Visualization of PMP routing displacement. Green: normal; Red: abnormal.

k-Shot	Method	MVTec-3D-AD			Eyecandies		
		AUROC$_I$	AUROC$_P$	AUPRO	AUROC$_I$	AUROC$_P$	AUPRO
5	BTF [10]	0.671	0.980	0.920	0.652	0.815	0.738
	AST [32]	0.680	0.950	0.903	0.633	0.741	0.691
	M3DM [41]	0.822	0.984	0.937	0.764	0.871	0.807
	CFM [6]	0.811	0.986	0.949	0.795	0.879	0.801
	3D-ADNAS [22]	0.826	-	-	0.775	0.875	-
	FIND [19]	0.899	0.991	**0.963**	0.840	0.951	0.862
	GPFlow	**0.912**	**0.992**	**0.963**	**0.913**	**0.982**	**0.905**
10	BTF [10]	0.695	0.983	0.928	0.685	0.834	0.806
	AST [32]	0.689	0.946	0.835	0.671	0.767	0.624
	M3DM [41]	0.845	0.986	0.943	0.824	0.890	0.812
	CFM [6]	0.845	0.987	0.954	0.838	0.903	0.825
	3D-ADNAS [22]	0.848	-	-	0.807	0.869	-
	FIND [19]	0.921	0.993	0.967	0.868	0.959	0.889
	GPFlow	**0.935**	**0.994**	**0.968**	**0.923**	**0.983**	**0.908**
50	BTF [10]	0.806	0.989	0.947	0.721	0.856	0.824
	AST [32]	0.794	0.974	0.929	0.739	0.862	0.715
	M3DM [41]	0.907	0.989	0.955	0.836	0.933	0.846
	CFM [6]	0.906	0.991	0.965	0.852	0.926	0.851
	3D-ADNAS [22]	0.890	-	-	0.868	0.912	-
	FIND [19]	0.952	**0.995**	**0.973**	0.897	0.976	0.907
	GPFlow	**0.955**	**0.995**	0.971	**0.925**	**0.985**	**0.913**
ALL	BTF [10]	0.865	0.992	0.959	0.740	0.883	0.845
	AST [32]	0.937	0.976	0.944	0.780	0.902	0.744
	M3DM [41]	0.945	0.992	0.964	0.882	0.977	0.887
	CFM [6]	0.954	0.993	0.971	0.881	0.974	0.887
	3D-ADNAS [22]	0.951	-	-	0.946	0.970	-
	FIND [19]	**0.978**	**0.995**	**0.983**	**0.955**	0.983	**0.925**
	GPFlow	0.970	**0.995**	0.976	0.942	**0.989**	**0.931**

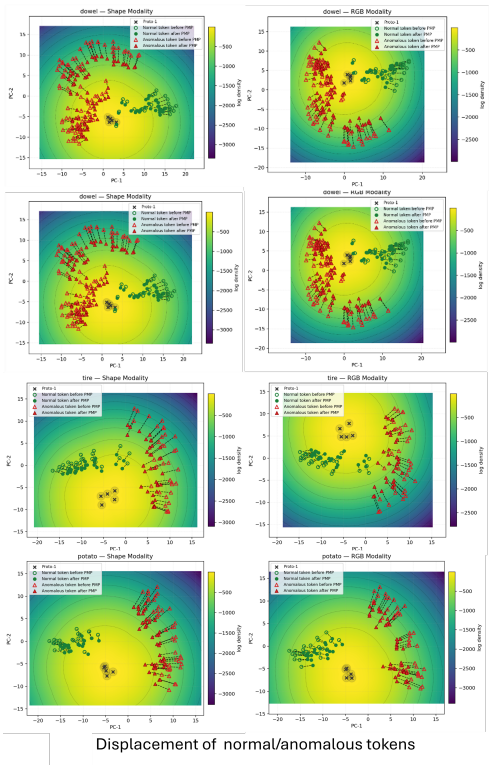

Table 2. Ablation on each component.

Method	AUROC$_I$	AUROC$_P$	AUPRO	$F_{1,\max,I}$
GPFlow w/o PMP	0.883	0.989	0.956	0.937
GPFlow w/o PCR	0.915	0.991	0.958	0.947
GPFlow w/o IPC	0.925	0.993	0.963	0.950
GPFlow	0.935	0.994	0.968	0.952

Table 3. Ablation on Gaussian prototypes & PMP router. OT: Optimal transport; SA: Soft assignment; SB: Schrödinger Bridge.

Method	AUROC$_I$	AUROC$_P$	AUPRO	$F_{1,\max,I}$
VQ-Codebook + OT	0.751	0.973	0.891	0.906
VQ-Codebook + SA	0.885	0.991	0.956	0.937
DPDL: Gaussian + SB	0.908	0.992	0.961	0.940
Ours: Gaussian + PMP	0.935	0.994	0.968	0.952

ting.

Effectiveness of Core Components. Tab. 2 evaluates the impact of the three key modules.

- **PMP:** Replacing our PMP router with an attention-based prototype router [26] (i.e., aggregating prototypes via learned attention scores without using covariance information) leads to a significant drop in performance, with AUROC$_I$ decreasing from 0.935 to 0.883. This confirms that the covariance-aware Gaussian prototype probability flow is crucial for aggressively shrinking features toward normal modes and filtering out anomalous variations.
- **PCR:** We observe a significant performance drop without the cross-modal reconstruction: GPFlow w/o cross-modal reconstruction (intra-modal only) yields 0.915 AUROC$_I$, versus 0.935 with the coupled design. This highlights the importance of leveraging cross-modal consistency.
- **IPC:** Disabling IPC (i.e., freezing the prototypes at test time) degrades performance (0.925 AUROC$_I$ without IPC, vs. 0.935 with the full model). This confirms that updating the prototype distribution with confident normal patches is beneficial: it expands coverage to new normal variations that are unseen during training and reduces false positives.

In all cases, the full GPFlow achieves the best overall AUROC and AUPRO performance, validating that each proposed component contributes to the overall effectiveness.

Ablations on Gaussian Prototypes and PMP Router. We next evaluate our design choices for representing normal patterns and the mechanism for routing features toward prototypes. In Tab.3, we compare against four variants: (i) a discrete vector-quantized (VQ) codebook with optimal

Table 4. Effect of intra- and cross-modal coupling (intra-only vs. cross-only vs. coupled).

Method	AUROC$_I$	AUROC$_P$	AUPRO	$F_{1,\max,I}$
Intra RGB-only	0.829	0.974	0.915	0.920
Intra Shape-only	0.858	0.965	0.888	0.934
Intra (RGB + Shape)	0.915	0.991	0.958	0.936
Cross RGB-only	0.607	0.967	0.866	0.891
Cross Shape-only	0.782	0.983	0.932	0.916
Cross (RGB + Shape)	0.775	0.986	0.937	0.919
Coupled RGB-only	0.847	0.983	0.935	0.922
Coupled Shape-only	0.898	0.981	0.928	0.942
Ours: Coupled (RGB + Shape)	0.935	0.994	0.968	0.952

transport (OT) matching for reconstruction; (ii) a VQ codebook with soft assignment (SA), allowing a linear combination of nearest codes; (iii) learnable Gaussian prototypes with a learned diffusion process via a Schrödinger Bridge (SB); and (iv) our proposed Gaussian prototypes with the analytical PMP router.

Gaussian prototypes consistently outperform discrete codebooks (e.g., VQ with soft-assignment: 0.885 I-AUROC), confirming the importance of modeling continuous normal variation. Moreover, our PMP router (0.935) surpasses the Schrödinger-Bridge approach (0.908), demonstrating more robust anomaly detection.

Robustness of IPC to anomaly contamination and ρ. Tab. 5 analyzes our Instance-Aware Prior Calibration (IPC) under varying MSM selection ratios ρ and oracle contamination stress. Across all values of ρ, the anomaly leakage remains negligible, ensuring that the calibration is mainly driven by the normal context. The ratio of $\rho = 0.6$ achieves the optimal performance trade-off. Overly restricting ρ degrades calibration performance due to insufficient coverage of normal contexts rather than to anomaly leakage. Oracle stress tests highlight IPC's robustness to contamination from anomalies. Even when injected with up to 50% anomalous patches, IPC maintains robust performance and consistently outperforms the baseline model without IPC. Severe degradation occurs only with 100% contamination, when normal patches are completely unavailable.

Ablations on Coupled Reconstruction Pathways. Tab.4 analyzes the detailed contributions of the intra-modal, cross-modal, and coupled reconstruction pathways. Only using intra-modal reconstruction (Intra RGB + Shape) achieves 0.915 AUROC$_I$. In contrast, relying only on cross-modal reconstruction (Cross RGB + Shape) performs poorly (0.775 AUROC$_I$). Without modeling modality-specific normal patterns, cross-modal consistency alone is insufficient for MAD tasks. By modeling both modality-specific feature deviations and cross-modal inconsistencies, the proposed coupled architecture achieves the best performance (0.935 AUROC$_I$).

Computational Efficiency. Despite its strong performance, GPFlow remains computationally efficient. Tab. 6 compares its computation cost with baseline methods. Our method requires approximately 129 GFLOPs per forward pass on a 224×224 input, which uses only 18% of computation cost of FIND (728 GFLOPs). Although the parameter count of GPFlow is higher (265M), GPFlow processes an image in 69ms, slightly faster than FIND (76ms). Overall, GPFlow achieves state-of-the-art accuracy with substantially lower computational overhead, making it well-suited for real-time industrial inspection applications.

Table 5. **Oracle contamination stress (no MSM) forces the IPC to include 25%/50%/100% GT anomalous patches during prototype refinement. FNR@1%FPR** uses the threshold determined by normal test scores at FPR=1%. **DispRatio**: ratio between the mean PMP displacement of normal/anomaly patches in test images (lower is better). **Leak(%)**: Fraction of GT anomaly patches included in the MSM candidate set.

Setting	I-AUROC↑	FNR@1%FPR↓	DispRatio↓	Leak(%)↓
No-IPC	0.894	0.272	0.798	–
IPC + MSM (top-ρ=0.2)	0.901	0.220	0.775	0.27
IPC + MSM (top-ρ=0.4)	0.908	0.179	0.769	0.75
IPC + MSM (top-ρ=0.6)	**0.911**	**0.151**	**0.763**	1.29
IPC + MSM (top-ρ=0.8)	0.907	0.156	0.761	1.98
Stress level (25% anomaly patches)	0.898	0.211	0.786	–
Stress level (50% anomaly patches)	0.895	0.234	0.799	–
Stress level (100% anomaly patches)	0.872	0.355	0.829	–

Table 6. Computation efficiency comparison. FLOPs are per forward pass at 224×224; latency measured with batch$= 1$.

Method	AUROC$_I$ ↑	FLOPs↓ (G)	#Params↓ (M)	Latency↓ (ms, b=1)
CFM	0.845	228.53	146.80	35.57
FIND	0.921	728.46	186.89	76.09
GPFlow	0.935	129.27	265.44	69.02

5. Conclusion

We introduced GPFlow, a framework for multi-modal anomaly detection. By modeling normality with learnable Gaussian priors and guiding feature reconstruction through an analytical Posterior-Mean Path router, GPFlow captures continuous normal variations while suppressing shortcut reconstructions through a covariance-aware information bottleneck. In addition, GPFlow robustly exploits complementary multi-modal cues and adapts to unseen normal patterns through instance-aware prior calibration. Experiments on MVTec-3D-AD and Eyecandies show that GPFlow achieves remarkable performance with minimal normal data, making it a practical industrial solution.

Acknowledgements

This work is supported by the Agency for Science, Technology and Research (A*STAR) under its MTC Programmatic Funds (Grant No. M23L7b0021).

References

[1] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP*, pages 202–213, 2022. 6

[2] Luca Bonfiglioli, Marco Toschi, Davide Silvestri, Nicola Fioraio, and Daniele De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *16th Asian Conference on Computer Vision (ACCV)*, pages 459–475, 2022. 6

[3] Joan Bruna and Jiequn Han. Provable posterior sampling with denoising oracles via tilted transport. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 3

[4] Tri Cao, Jiawen Zhu, and Guansong Pang. Anomaly detection under distribution shift. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023. 2

[5] Yu-Min Chu, Chieh Liu, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3d anomaly detection. In *Proceedings of the 40th International Conference on Machine Learning (ICML)*, 2023. 2

[6] Alex Costanzino, Pierluigi Zama Ramirez, Giuseppe Lisanti, and Luigi Di Stefano. Multimodal industrial anomaly detection by crossmodal feature mapping. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17234–17243, 2024. 1, 2, 6, 7

[7] 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 (CVPR)*, 2022. 1

[8] Matic Fučka, Vitjan Zavrtanik, and Danijel Skočaj. Transfusion – a transparency-based diffusion model for anomaly detection. In *European Conference on Computer Vision (ECCV)*, 2024. 1

[9] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 6840–6851, 2020. 2, 3

[10] Eliahu Horwitz and Yedid Hoshen. Back to the feature: Classical 3D features are (almost) all you need for 3D anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 2968–2977, 2023. 6, 7

[11] Zahra Kadkhodaie and Eero P. Simoncelli. Stochastic solutions for linear inverse problems using the prior implicit in a denoiser. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 3

[12] Tero Karras, Miika Aittala, Timo Aila, and Samuli Laine. Elucidating the design space of diffusion-based generative models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 2

[13] Kwanyoung Kim and Jong Chul Ye. Noise2score: Tweedie's approach to self-supervised image denoising without clean images. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 3

[14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015. 6

[15] Branislav Kveton, Boris Oreshkin, Youngsuk Park, Aniket Deshmukh, and Rui Song. Online posterior sampling with a diffusion prior. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 2

[16] Joo Chan Lee, Taejune Kim, Eunbyung Park, Simon S. Woo, and Jong Hwan Ko. Continuous memory representation for anomaly detection. In *European Conference on Computer Vision (ECCV)*, 2024. 1

[17] Wenqiao Li, Bozhong Zheng, Xiaohao Xu, Jinye Gan, Fading Lu, Xiang Li, Na Ni, Zheng Tian, Xiaonan Huang, Shenghua Gao, and Yingna Wu. Multi-sensor object anomaly detection: Unifying appearance, geometry, and internal properties. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9984–9993, 2025. 1, 2

[18] Xurui Li, Ziming Huang, Feng Xue, and Yu Zhou. MuSc: Zero-shot industrial anomaly classification and segmentation with mutual scoring of the unlabeled images. In *The Twelfth International Conference on Learning Representations*, 2024. 5

[19] Yiting Li, Fayao Liu, Jingyi Liao, Sichao Tian, Chuan-Sheng Foo, and Xulei Yang. FIND: Few-shot anomaly inspection with normal-only multi-modal data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 23290–23299, 2025. 1, 2, 6, 7

[20] Yiting Li, Xulei Yang, Jing Zhang, Sichao Tian, Jingyi Liao, and Fayao Liu. PIRN: Prototypical-based intra-modal reconstruction with normality communication for multi-modal anomaly detection. In *The Fourteenth International Conference on Learning Representations*, 2026. 1, 2, 4

[21] Chen Liang, Wenguan Wang, Jiaxu Miao, and Yi Yang. GMMSeg: Gaussian mixture based generative semantic segmentation models. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 1

[22] Kaifang Long, Guoyang Xie, Lianbo Ma, Jiaqi Liu, and Zhichao Lu. Revisiting multimodal fusion for 3D anomaly detection from an architectural perspective. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12273–12281, 2025. 1, 2, 6, 7

[23] Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. DPM-Solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 2, 3

[24] 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*, 36:8487–8500, 2023. 1, 2, 4

[25] Yiwei Lu, Frank Yu, Mahesh Kumar Krishna Reddy, and Yang Wang. Few-shot scene-adaptive anomaly detection. In *European Conference on Computer Vision (ECCV)*, pages 125–141, 2020. 2

[26] Wei Luo, Yunkang Cao, Haiming Yao, Xiaotian Zhang, Jianan Lou, Yuqi Cheng, Weiming Shen, and Wenyong Yu.

Exploring intrinsic normal prototypes within a single image for universal anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 9974–9983, 2025. 1, 2, 4, 5, 6, 7

[27] Hila Manor and Tomer Michaeli. On the posterior distribution in denoising: Application to uncertainty quantification. In *The Twelfth International Conference on Learning Representations*, 2024. 3, 4

[28] Declan McIntosh and Alexandra Branzan Albu. Unsupervised, online and on-the-fly anomaly detection for non-stationary image distributions. In *European Conference on Computer Vision (ECCV)*, pages 428–445, 2024. 2

[29] Chenlin Meng, Yang Song, Wenzhe Li, and Stefano Ermon. Estimating high order gradients of the data distribution by denoising. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021. 3

[30] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy V. Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. DINOv2: Learning robust visual features without supervision. *Transactions on Machine Learning Research*, 2024. 6

[31] Marien Renaud, Jiaming Liu, Valentin de Bortoli, Andrés Almansa, and Ulugbek S. Kamilov. Plug-and-play posterior sampling under mismatched measurement and prior models. In *International Conference on Learning Representations (ICLR)*, 2024. 3

[32] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV*, pages 2592–2602, 2023. 6, 7

[33] Kulin Shah, Sitan Chen, and Adam R. Klivans. Learning mixtures of gaussians using the ddpm objective. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023. 2

[34] Andrew H. Song, Richard J. Chen, Tong Ding, Drew F.K. Williamson, Guillaume Jaume, and Faisal Mahmood. Morphological prototyping for unsupervised slide representation learning in computational pathology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11566–11578, 2024. 1

[35] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations (ICLR)*, 2021. 2

[36] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 11895–11907, 2019. 3

[37] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021. 2, 3

[38] Yuanpeng Tu, Boshen Zhang, Liang Liu, Yuxi Li, Jiangning Zhang, Yabiao Wang, Chengjie Wang, and Cairong Zhao. Self-supervised feature adaptation for 3d industrial anomaly detection. In *European Conference on Computer Vision (ECCV)*, pages 75–91. Springer, 2024. 2

[39] Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011. 2, 3

[40] Fuyun Wang, Tong Zhang, Yuanzhi Wang, Yide Qiu, Xin Liu, Xu Guo, and Zhen Cui. Distribution prototype diffusion learning for open-set supervised anomaly detection. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 20416–20426, 2025. 1, 2

[41] Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8032–8041, 2023. 1, 2, 6, 7

[42] Ze Wang, Yipin Zhou, Rui Wang, Tsung-Yu Lin, Ashish Shah, and Ser Nam Lim. Few-shot fast-adaptive anomaly detection. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. 2

[43] Bingchen Zhao, Xin Wen, and Kai Han. Learning semi-supervised gaussian mixture models for generalized category discovery. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 16623–16633, 2023. 1

[44] Zheyuan Zhou, Le Wang, Naiyu Fang, Zili Wang, Lemiao Qiu, and Shuyou Zhang. R3D-AD: Reconstruction via diffusion for 3d anomaly detection. In *European Conference on Computer Vision (ECCV)*, 2024. 1

[45] Wenbing Zhu, Lidong Wang, Ziqing Zhou, Chengjie Wang, Yurui Pan, Ruoyi Zhang, Zhuhao Chen, Linjie Cheng, Bin-Bin Gao, Jiangning Zhang, Zhenye Gan, Yuxie Wang, Yulong Chen, Shuguang Qian, Mingmin Chi, Bo Peng, and Lizhuang Ma. Real-IAD D3: A real-world 2d/pseudo-3d/3d dataset for industrial anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15214–15223, 2025. 1, 2