MegAD: An Expert in Meta-learning Guided Few-shot Anomaly Detection

Supplementary Material

1. Prompt template

As shown in Fig. 1, the normal and basic prompt templates on which MegAD is based are derived from AnomalyGPT [16]. The text consists of templates, status words, and class names. If 'a photo of a ' is used to represent the basic template, the category where the data set is located is in , and the category of each sample is known. Therefore, we can obtain the corresponding prompt statements to guide the model learning through the given template and the current normal state. In addition, we have provided five example anomaly prompt suffixes generated by MPG for each dataset, as detailed in Table 1.

2. Preliminaries

The Mixture of Experts (MoE) framework consists of a set of specialized expert models and a differentiable gating network. MoE aims to handle complex tasks by decomposing the input space into subspaces, where each expert focuses on modeling specific patterns, and the gating network dynamically coordinates their contributions. Formally, given a set of N parameterized expert functions $\{E_i(x)\}_{i=1}^{N}$, each expert specializes in distinct regions of the input space. The gating network $G(x)$ generates a probability distribution over the experts, reflecting their relevance to the input x. The final output y is computed as the weighted sum of the expert predictions:

$$y = \sum_{i=1}^{N} G_i(x) \cdot E_i(x), \tag{1}$$

where $G_i(x)$ denotes the gating weight for the i-th expert, satisfying $\sum_{i=1}^{N} G_i(x) = 1$. However, a critical challenge in MoE is the *expert dominance* problem, where a subset of experts monopolizes the predictions, leading to underutilization of the remaining experts. To address this, prior work introduces an auxiliary loss term to encourage balanced expert utilization. Let P_i represent the importance weight of the i-th expert, computed as the batch-averaged gating probability:

$$P_i = \frac{1}{B} \sum_{b=1}^{B} G_i(x_b), \tag{2}$$

where B is the batch size. The load-balancing loss minimizes the Kullback-Leibler (KL) divergence between P and a uniform distribution U:

$$\mathcal{L}_{\text{balance}} = \lambda_{\text{bal}} \cdot \text{KL}(P \parallel U) = \lambda_{\text{bal}} \cdot \left(\sum_{i=1}^{N} P_i \log \frac{P_i}{1/N} \right), \tag{3}$$

where λ_{bal} is a balancing coefficient. This loss penalizes skewed expert utilization, ensuring all experts contribute meaningfully to the predictions.

3. Pseudocode Descriptions

In this section, we will provide pseudocode for the proposed modules: Dynamic Assistance Mixture of Experts (DA-MOE), Meta-Learning Guided Anomaly Generator (MPG), and Graph Enhanced Component Modeling (M-Former). These pseudocode representations are intended to give readers a clearer and more structured understanding of the implementation details of each module. Algorithm 1, Algorithm 2, and Algorithm 3 illustrate the pseudocode for these three modules, respectively. In the DA-MOE module, a dynamic assist learning mechanism is employed to optimize the collaborative capability of the mixture of experts' architecture. Specifically, by calculating the confidence scores of the expert models, high-performing and low-performing experts are dynamically assigned as helpers and learners, respectively. The interaction between experts is based on knowledge distillation loss, parameter similarity constraints, and gradient alignment strategies to achieve consistency in cross-expert semantic modeling. Simultaneously, by integrating global semantic features and real-time confidence scores through a dynamic gating network, the participation weights of low-confidence experts are explicitly boosted, thereby balancing the coverage of multi-class anomaly features and alleviating inter-class disparity issues.

The VAE module comprises an encoder E_ψ and decoder D_ϕ. E_ψ maps P_c to parameters of a Gaussian distribution in latent space:

$$\mu, \log \sigma^2 = E_\psi(P_c), \quad z \sim \mathcal{N}(\mu, \sigma^2) I \tag{4}$$

where $z \in \mathbb{R}^d$ is the latent code. The decoder D_ϕ reconstructs anomaly suffix embeddings A_c from z:

$$A_c = D_\phi(z) = \text{MLP}(\text{LayerNorm}(zW_z)) \tag{5}$$

with W_z as a learnable projection. To further promote the use of MPG, during the inference phase, A_c is synthesized by sampling z from the learned posterior $q(z|P_c)$, ensuring adaptive generation for unseen classes. The final text features $F_{\text{text}} = [F_n, F_a]$ are derived by encoding fixed normal prompts and A_c via ImageBind encoder.

Normal templates:

```
prompt_normal = ['{}', 'flawless {}', 'perfect {}', 'unblemished {}', '{} without flaw', '{} without defect', '{} without damage']
```

Basic templates:

```
prompt_templates = ['a photo of a {}.', 'a photo of the {}.']
```

Figure 1. List of normal and basic text prompts.

Table 1. Anomaly Suffix for Different Datasets

Datasets	Generate anomaly suffix
MVTec-AD	' with large breakage', 'with bent wire and missing insulation', 'with color stain and chipped coating', 'with scratch head and deformed thread', 'with broken teeth and fabric detachment'
MVTec-AD2	' with internal crack under dark field', 'with misaligned teeth and metal shavings', 'with solder bridging under 10x magnification', ' with laser-induced microfractures', ' with thermal deformation in vacuum environment'
VisA	' with melted wax anomaly and embedded foreign particles', 'with seal failure and content leakage', 'with structural deformation and surface discoloration', 'with abnormal cement distribution', ' with thermal stress cracks'
DAGM	' with elliptical thread break pattern', ' with oil contamination spreading radially', ' with irregular warp distortion', ' with cross-weave contamination', ' with diagonal yarn rupture'
MPDD	' with stress corrosion cracking', ' with pitting corrosion under joint', ' with contact pad delamination', ' with fatigue cracks near bolt holes', ' with incomplete penetration'
MulSen-AD	' with micro-cracks and hot spots', ' with gas inclusion and wall thinning', ' with torsion failure and metal fatigue', ' with synthetic fiber contamination', ' with solder ball bridging'
BrainMRI	' with irregular gadolinium enhancement', ' with leptomeningeal metastasis', ' with necrotic core and peritumoral edema', 'with cortical dysplasia', ' with abnormal white matter hyperintensities'
LiverCT	' with hypodense metastatic lesion', 'with portal venous gas', ' with irregular perfusion defect', ' with intraductal calcifications', ' with traumatic laceration'
ReSC	' with cystoid macular edema', ' with subretinal fluid accumulation', ' with drusen clustering', ' with ellipsoid zone disruption', ' with vitreomacular traction'

4. More Discussions

4.1. Meta learning

The core goal of meta-learning [14] is to enable models to learn new tasks from a small number of samples quickly. Many previous works [13, 48] have shown that meta-learning can effectively address FSAD problems by transferring prior knowledge from historical tasks, and it exhibits good anti-overfitting characteristics when adapting to new tasks. Recent work [] has further expanded the application boundaries of meta-learning. For instance, ACR [35] integrates meta-learning strategies with batch normalization mechanisms to construct an efficient and general anomaly detection framework. MetaUAS [15] proposes single-prompt meta-learning, which enables rapid adaptation to new tasks with just a single normal sample, providing a methodological foundation for the deep integration of meta-learning and prompt learning. However, previous work still has many shortcomings. Therefore, this paper introduces a meta-learning guided anomaly prompt generation module that can adaptively generate context-aware anomaly patterns. This module addresses the high time costs associated with tradi-

Algorithm 1 DA-MOE

Input: Training batch $\mathcal{D}_{batch}$, global semantic features F_{cls}, set of K expert networks $\{E_1, E_2, \ldots, E_K\}$, alignment strength α, boosting scale γ, initial decay parameters λ_1^0, λ_2^0

Output: Updated expert parameters θ, gating network weights W_g

for each training batch $x \in \mathcal{D}_{batch}$ **do**
  for each expert $E_k \in \{E_1, \ldots, E_K\}$ **do**
    Calculate confidence score $\mathcal{C}_k$ at current step using Eq. 1.
  end for
  Compile confidence scores $s = [\mathcal{C}_1, \ldots, \mathcal{C}_K]$
  Rank experts by $\mathcal{C}_k$ in descending order
  Designate top-k experts as Helpers (E_{Helper}) and bottom-k as Learners ($E_{Learner}$)
  Calculate Knowledge Transfer Guided Loss $\mathcal{L}_{KTG}$ using Eq. 2.
  Compute dynamic gating weights g using Eq. 3.
  Update parameters of experts and gating network via joint optimization.
  Decay λ_1, λ_2 exponentially: $\lambda_i(t) = \lambda_i^0 \cdot e^{-\eta t}$.
end for

Algorithm 2 MPG

Input: Normal support visual features $\{V_i^s\}$, fixed normal text prompts, training tasks $\mathcal{T}_i$, VAE encoder E_ψ, VAE decoder D_ϕ

Output: Anomaly suffix embeddings A_c, final cross-modal text features F_{text}, updated VAE parameters ψ, ϕ

for each training task $\mathcal{T}_i$ **do**
  Aggregate normal support features into class prototype P_c by Eq. 4.
  Map P_c to Gaussian distribution parameters in latent space by Supp. Eq. 4.
  Sample latent variable z using the reparameterization trick $z \sim \mathcal{N}(\mu, \sigma^2 I)$.
  Reconstruct diverse anomaly suffix embeddings A_c by Supp. Eq. 5.
  Encode fixed normal prompts and A_c via ImageBind encoder to derive text features $F_{text} = [F_n, F_a]$.
  Calculate meta-learning loss $\mathcal{L}_{meta}$ by Eq. 5.
  Update parameters ψ and ϕ to optimize the mapping from latent space to text space.
end for

tional methods that rely on manually designed text prompts, enabling flexible and adaptive anomaly detection, particularly enhancing the model's robustness in scenarios with few samples and complex defects.

Algorithm 3 M-Former

Input: Multi-layer visual features $\{V_i\}$, text features $F_{text} = [F_n, F_a]$, temperature parameter τ, balancing weight γ, losses $\mathcal{L}_{KTG}$ and $\mathcal{L}_{meta}$

Output: Cross-Modal Contrastive Alignment loss $\mathcal{L}_{CMCA}$, Total training loss $\mathcal{L}_{total}$

for each visual feature layer V_i **do**
  Compute channel-wise variance $Var(V_i)$ to capture potential anomalies.
  Calculate spatial attention mask A_i highlighting high variance regions by Eq. 6.
  Generate dynamic anomaly query vectors Q_i using A_i and V_i by Eq. 7.
  Compute Bidirectional Contrastive Loss $\mathcal{L}_{CMCA}$ to align Q_i with F_n and F_a by Eq. 8.
end for
Calculate total training loss $\mathcal{L}_{total} = \mathcal{L}_{KTG} + \mathcal{L}_{meta} + \gamma\mathcal{L}_{CMCA}$ by Eq. 9.

4.2. Few-shot selection strategy

In this paper, we adopt a few-shot strategy for anomaly detection by constructing a few-shot reference library through random selection and fixation of normal samples. Although random selection avoids human-induced bias and simplifies experimental workflows, its inherent statistical representativeness issues may raise concerns about the rationality of randomly chosen samples and potential inadequacies in their impact on test results. To validate the feasibility of the random strategy, we compared the sensitivity of different baseline models (e.g. SPADE [10], PatchCore [45]) to random samples under a unified experimental setup. Empirical results demonstrate that random selection is viable as long as all methods utilize identical reference samples, ensuring fair result comparisons. However, further analysis revealed significant discrepancies in outcomes across different hardware configurations (e.g. GPUs, CUDA versions), stemming from uncontrollable perturbations in the parallel computing mechanisms of deep learning frameworks. To address this, we ultimately fixed the selected samples in our experiments, enabling readers to reproduce our results within an acceptable error margin. This approach balances methodological rigor with practical reproducibility in resource-constrained industrial scenarios.

In practical applications, implementing a random few-shot strategy is challenging and cannot achieve the rigor of the 1-shot, 2-shot, and 4-shot settings described in this paper. A feasible approach is to use K-means clustering to select prototype samples from the training set that cover all normal patterns, while ensuring distributional consistency via the LPIPS metric, thereby enabling unrestricted applicability in real-world scenarios.

4.3. Model complexity

Model parameters and computational costs are critical metrics for evaluating models. Therefore, we assessed the parameter count (M) and inference speed (frames per second, FPS) of the baseline models and our proposed model. To ensure fair comparisons, all experiments were conducted on a single NVIDIA RTX 4090 24GB GPU without interference from other running processes. As shown in Tab. 2, under identical experimental settings and a fixed image size of 224×224, RegAD [23] achieves the smallest parameter count (25.2M) and the fastest inference speed. While MegAD has a larger parameter size, its inference speed remains comparable, with the advantage of stronger performance.

5. Datasets

MVTec-AD The MVTec-AD [4] dataset is widely used as a standard benchmark for evaluating unsupervised anomaly detection methods. This dataset contains 5354 high-resolution images (3629 images for training and 1725 images for testing) of 15 different product categories. 5 classes consist of textures and the other 10 classes contain objects. A total of 73 different defect types are presented and almost 1900 defective regions are manually annotated in this dataset.

MVTec-AD2 The MVTec-AD2 [21] is a dataset that includes 8 types of high-difficulty detection scenarios (such as transparent/overlapping object defects, dark field illumination interference, and small defect localization), containing 379 normal samples and 705 abnormal samples.

VisA The VisA [57] dataset is a larger anomaly detection dataset compared to MVTecAD [4]. This dataset contains 10821 images with 9621 normal and 1200 anomalous samples. In addition to images that only contain single instance, the VisA dataset also have images that contain multiple instances. Moreover, some product categories of the VisA dataset, such as Cashew, Chewing gum, Fryum and Pipe fryum, have objects that are roughly aligned. These characteristics make the VisA dataset more challenging than the MVTecAD dataset, whose images only have single instance and are better aligned.

MPDD The MPDD [25] contains 6 classes of metal parts, comprising 888 normal samples for the training set and 458 samples either normal or anomalous in the test set. Because of the variable spatial orientation, position, and distance of multiple objects concerning the camera at different light intensities and with a non-homogeneous background, this dataset is a more challenging dataset.

DAGM The DAGM [49] dataset includes 10 types of textured surface defect images, comprising 6,996 normal images for the training set and 1,054 abnormal images for the test set. It was originally created for a competition at the 2007 symposium of the DAGM (Deutsche Arbeitsgemeinschaft für Mustererkennung e.V., the German chapter of the International Association for Pattern Recognition).

MulSen-AD The MulSen-AD [37] dataset is the first multi-sensor industrial anomaly detection dataset, providing RGB images, 3D laser point clouds, and infrared thermal imaging data simultaneously (this paper only uses the RGB modality). It covers 15 types of real industrial product anomalies, with the training set containing 150 normal samples and the test set containing 360 abnormal samples.

BrainMRI The BrainMRI [2] dataset is based on the BraTS2021 [1] dataset, one of the latest large-scale brain tumor segmenta- tion datasets. It contains complete 3D brain volume images. The BrainMRI dataset consists of 2D slices derived from BraTS2021, with each slice image measuring 240×240 pixels. The training set includes 7,500 normal samples, and the test set contains 3,715 samples, both normal and anomalous, with pixel-level anomaly annotations.

LiverCT The LiverCT [2] dataset is constructed from the BTCV [34] and LiTS [6] datasets. It contains 50 normal abdominal 3D CT scans from BTCV and 131 abdominal 3D CT scans, both normal and anomalous, from LiTS. The Hounsfield Unit (HU) values of the 3D scans from both datasets are con- verted to grayscale using the abdominal window and then cropped into 2D slices. The dataset includes 1,452 normal 2D slices for training and 1,493 2D slices, both normal and anomalous, for testing, with a resolution of 512×512 and pixel-level anomaly annotations

RESC The Retinal Edema Segmentation Challenge (RESC) [22] dataset is a retinal OCT dataset containing 4,297 normal images for training and 1,805 test images, both nor- mal and anomalous. The image resolution is $512 \times 1,024$, and the dataset provides pixel-level anomaly annotations.

6. Additional ablation Results

Impact of hyperparameters on Mixture Of Experts. The number of experts (E) and the Top k coefficient (K) together determine the balance between model capacity and generalization ability, so the choice of hyperparameters (E, K) has a significant impact on the performance of the MOE module. As shown in Fig. 2, when we set E=5 and K=2, the model achieves optimal performance in the anomaly detection task (96.7/97.2 for I-AUROC/P-AUROC). At this point, the expert collaboration mechanism effectively enhances the accuracy of anomaly localization and classification. This

Table 2. Model complexity comparison between our MegAD and other competing methods.

	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**MegAD**
Parameters(M)	74.5	686.9	69.5	25.2	165.9	662.3	411.5	442.6	482.7	457.2
FPS	4.8	14.1	21.5	20.2	0.51	17.2	15.4	18.8	18.2	17.9

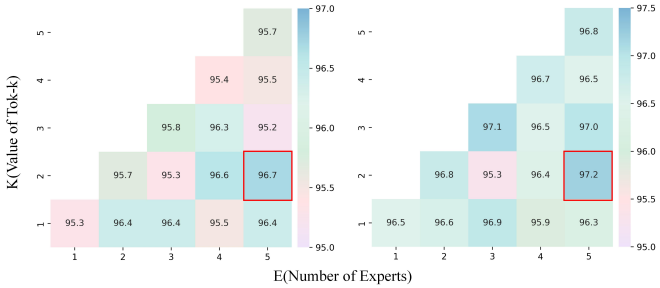


Figure 2. I-AUROC (left) and P-AUROC (right) of different number of experts E and value of Top-K K on the generalization performance, calculated over MVTec-AD dataset.

Table 3. Ablation study on the threshold parameter τ.

τ	MVTec-AD		VisA	
	I-AUROC	P-AUROC	I-AUROC	P-AUROC
1	96.3	96.8	91.8	97.1
2	96.7	97.2	92.4	97.6
3	96.5	97.1	92.1	97.4
4	96.2	96.8	91.5	97.0

parameter configuration achieved precise localization and reliable classification of sample anomalies on the MVTec-AD dataset.

Threshold Parameter Ablation. To analyze the impact of threshold τ on model performance, we conducted systematic ablation experiments. As shown in Table 3, the model achieves optimal performance on both MVTec-AD and VisA datasets when $\tau = 2$. Performance follows an inverted U-shaped pattern as τ increases. When $\tau > 2$, the average I-AUROC on MVTec-AD and VisA decreases by 0.2%, indicating that excessive expansion of Helper-Learner pairs can weaken the expert anomaly detection characteristics. Therefore, we adopt $\tau = 2$ in all experiments to balance knowledge transfer intensity with model discriminative power.

MPG ablation. Our proposed MPG module replaces manual prompt engineering with a variational meta-learning framework, which dynamically generates anomaly prompts from normal support features. This enables the model to map normal features to anomaly descriptions through latent space sampling $z \sim \mathcal{N}(\mu, \sigma^2)$. The MPG synthesizes context-aware prompts, making it suitable for unseen anomalies, thereby enhancing the model's generalization capability (e.g., transitioning from industrial domains to medical domains). As shown in Table 5, we conducted separate ablation experiments on MPG using MVTec-AD2 and MulSen-AD to obtain a more comprehensive analysis.

Table 4. Ablation study on the M-Former components. Removing the Dynamic Anomaly Query Generation (DAQG) module primarily degrades image-level classification (I-AUROC), while disabling the Cross-Modal Context Alignment (CMCA) loss more significantly impacts pixel-level localization (P-AUROC), validating their distinct and complementary roles.

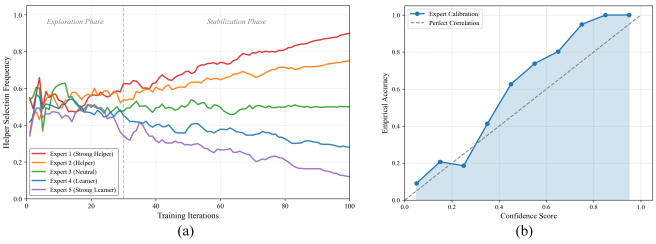
M-Former	MVTec-AD		VisA	
	I-AUROC	P-AUROC	I-AUROC	P-AUROC
w/o DAQG	96.1	97.1	91.5	97.2
w/o CMCA	96.8	96.5	92.2	96.9
both	96.7	97.2	92.4	97.6

M-Former ablation. We systematically decoupling its two key components – the DAQG and the CMCA, and independently evaluating them. As shown in the Table 4, removing DAQG from the MVTec-AD dataset resulted in a significant decrease of 0.6% in image-level anomaly detection performance, while removing CMCA resulted in a decrease of 0.7% in pixel-level localization accuracy. This downward trend in performance was more pronounced on the cross-domain dataset VisA. Experimental analysis shows that DAQG plays a decisive role in improving the semantic-level anomaly classification performance through its multi-scale dynamic focusing mechanism (Eq. 7), while the CMCA loss significantly improves the model's fine-grained localization ability by forcing the semantic alignment of visual features and text prompts (Eq. 8).

In addition, integrating our M-Former module with the AnomalyCLIP baseline demonstrates that M-Former significantly enhances cross-modal alignment performance. We attribute this improvement to M-Former's ability to precisely capture local anomalous features through channel-variance attention, and its bidirectional contrastive alignment loss compresses the standard deviation of visual-textual feature similarity compared to the baseline. For CLIP-based models, the inherent global feature fusion through addition struggles to model local anomaly semantics under limited model depth. In contrast, we observe that DAQG's Q_i adaptively concentrates on anomalous regions of varying sizes and shapes,

Table 5. Ablation study of the Meta Prompt Generator (MPG). Replacing fixed prompt templates with our variational meta-learning framework consistently improves both image- and pixel-level anomaly detection across domains, demonstrating enhanced generalization capability.

Prompt Method	MVTec-AD2		MulSen-AD	
	I-AUROC	P-AUROC	I-AUROC	P-AUROC
Fixed Templates	64.2	80.4	86.1	95.0
MPG (Ours)	66.6	82.9	89.9	95.6



(a) (b)

Figure 3. Stability and reliability of AL mechanism experiments.

enabling AnomalyCLIP to prioritize critical areas during cross-modal alignment.

Heuristic Nature of Assisted Learning. We conducted a correlation analysis on the MVTec-AD validation set. To address concerns regarding potential instability in role assignment, we performed a convergence analysis of the dynamic changes in the "assistant/learner" roles during training. As shown in Fig. 3(a), role assignments exhibited considerable fluctuations during the early "exploration" phase. However, as training progressed, the curves stabilized significantly and gradually diverged. We calculated the variance of role selection frequency in the final 10% of iterations, which was found to be extremely low ($\sigma^2 < 2 \times 10^{-4}$), indicating that the experts had stably specialized in distinct feature patterns without oscillations. This validates the convergence of the auxiliary learning mechanism. Furthermore, as illustrated in Fig. 3(b), the Spearman correlation coefficient between experts' prediction confidence and their task accuracy reached 0.55, demonstrating a strong correlation and indicating that experts with higher confidence statistically made significantly more accurate predictions.

Loss ablation. We conduct incremental ablation studies to evaluate the contributions of the proposed loss functions, as shown in Tab. 6. When only the L_{KPG} loss is enabled, the model achieves 96.0 I-AUC and 96.3 P-AUC on MVTec-AD, 91.8 I-AUC and 97.1 P-AUC on VisA. Adding both L_{meta} and L_{CMCA} losses improves performance to 96.7 I-AUC (2.2 ↑) and 96.9 P-AUC on MVTec-AD, and 92.2 I-AUC (0.4 ↑) and 97.5 P-AUC on VisA. Notably, the full

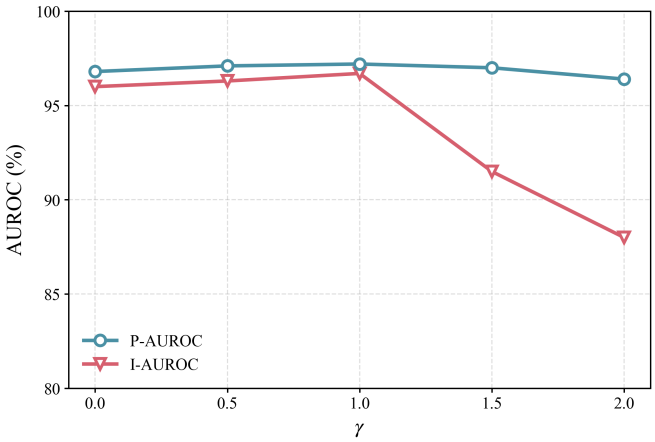


Figure 4. Ablation on the fusion coefficient γ at the 1-shot setting on the MVTec-AD dataset.

configuration with all three losses achieves the highest results: 96.7% I-AUC, 97.2% P-AUC, and 93.6% accuracy on MVTec-AD, and 92.4% I-AUC, 97.6% P-AUC, and 88.0% accuracy on VisA. This demonstrates that L_{KPG} provides a strong baseline, while L_{meta} and L_{CMCA} synergistically enhance cross-scale feature alignment and meta-learning robustness. The progressive improvements across metrics confirm the necessity of integrating all three components for optimal anomaly detection and localization in industrial inspection scenarios.

Loss hyperparameter analysis. As shown in Fig. 4, we conducted a comprehensive ablation study on the fusion coefficient γ in the overall loss function, evaluating its impact under one-shot settings on the MVTec-AD benchmark. We varied γ from 0 to 2.0, with the best results achieved when $\gamma = 1$. Therefore, we justify our choice of the default value $\gamma = 1.0$ for all experiments, which achieves the optimal trade-off between multimodal consistency and anomaly distinguishability.

The influence of different sample selections. In the main text, we take the average of five different random sampling results as our outcome. However, the impact of sample selection on detection performance remains a critical focus for readers. Therefore, Tab. 7 presents the results of five random sample groups under the 1-shot setting across different datasets.

7. Experiments on more challenging settings

We conducted further validation under more challenging experimental settings. Since the VisA dataset does not contain multi-sensor data images like MulSen-AD, we set up training on the VisA dataset and testing on MulSen-AD, analogous to

Table 6. Loss function ablation. Performance improves progressively as L_{meta} and L_{CMCA} are added to the L_{KPG} baseline, confirming their complementary roles in achieving state-of-the-art results.

L_{KPG}	L_{meta}	L_{CMCA}	MVTec-AD			VisA		
			I-AUC	P-AUC	Acc	I-AUC	P-AUC	Acc
✓			96.0	96.3	92.8	91.8	97.1	87.3
✓	✓		96.7	96.9	93.4	92.2	97.5	87.6
✓	✓	✓	96.7	97.2	93.6	92.4	97.6	88.0

Table 7. Impact of sample selection on anomaly detection performance. Each entry shows the metric triplet (A/B/C) for five different random samples under the 1-shot setting. The final column reproduces the averaged results from Tab. 1, confirming the minimal variance induced by sample randomness.

Domain	Dataset	Sample1	Sample2	Sample3	Sample4	Sample5	Result in Tab. 1
Industrial	MVTec AD	96.5/93.4/97.0	97.0/93.8/97.5	96.0/93.2/96.8	97.3/93.9/97.6	96.9/93.7/97.3	96.7/93.6/97.2
	MVTec AD2	66.0/69.0/82.0	67.0/69.8/83.0	66.5/69.4/82.5	66.3/69.3/82.8	67.3/69.9/83.2	66.6/69.5/82.9
	VisA	92.0/87.5/97.0	92.5/88.0/97.5	92.2/87.8/97.3	92.7/88.2/97.7	92.3/87.5/97.1	92.4/88.0/97.6
	DAGM	98.0/97.0/94.3	98.4/97.5/94.3	98.3/97.2/94.3	98.5/97.6/94.3	98.2/97.5/94.1	98.4/97.4/94.2
	MPDD	76.0/78.0/97.0	77.0/79.0/97.5	76.5/78.5/97.2	76.8/78.8/97.3	77.2/79.2/97.7	76.9/78.9/97.2
	MulSen-AD	89.2/86.5/94.8	90.1/87.5/96.6	89.3/87.0/95.7	90.5/87.7/96.2	89.7/87.2/95.2	89.9/87.4/95.6
Medical	BrainMRI	73.0/90.0/96.0	75.0/91.0/97.0	73.5/90.5/96.5	74.5/91.5/97.5	74.0/90.9/96.8	74.0/90.9/96.5
	LiverCT	63.0/64.0/98.0	63.5/64.5/98.5	63.2/64.2/98.2	63.7/64.7/98.7	63.6/64.4/98.4	63.4/64.3/98.3
	ReSC	86.3/78.0/96.6	87.1/79.0/97.8	86.3/78.5/96.1	86.4/78.8/96.6	86.0/78.2/96.4	86.3/78.4/96.8

Table 8. Performance comparison under the challenging quasi-zero-shot setting (training on VisA, testing on MulSen-AD).

Method	I-AUROC	I-Acc	P-AUROC	P-F1	P-AP	P-PRO
AnomalyGPT	83.0	86.4	96.2	53.8	87.3	89.4
MegAD	89.9	93.6	95.6	59.5	91.0	93.2
MegAD †	90.0	92.8	96.7	58.9	91.3	93.7

zero-shot anomaly detection. As shown in the Table 8, under this quasi-zero-shot setting, compared to the original setup, the performance metrics decreased by only 0.8%. Contrasted with baseline models, it still maintains strong prompt adaptation capabilities and effectively alleviates ICD. Beyond the SOTA performance of our model under the standard settings shown in Table 1 of the main text, the new experiments more prominently demonstrate MegAD's core strengths in alleviating ICD and adapting prompts.

8. Detailed Comparison Results

In this section, we provide a detailed comparison of the 1-shot performance across different models on the datasets adopted in this paper. Additionally, we evaluate and present the image-level accuracy (Accuracy) performance. Specifically, the results on the MVTec AD are shown in Tab. 9– 11, the results on the MVTec AD 2 are shown in Tab. 12– 14, the results on the VisA are shown in Tab. 15- 17, the results on the DAGM are shown in Tab. 18– 20, the results on the MPDD are shown in Tab. 21– 23, the results on the MulSen-AD are shown in Tab. 24– 26, and the medical datasets are shown in Tab, 27- 29. MegAD achieves state-of-the-art (SOTA) performance across nearly all metrics. While PromptAD [38] performs relatively well on certain datasets, it is constrained by manually crafted prompt templates for few-shot training, and the baseline CLIP lacks prior knowledge in anomaly detection. In contrast, our method generates meta-learned anomaly prompts without requiring manual template design, enabling robust anomaly detection across diverse datasets.

9. Additional Qualitative Results.

We have visualized additional qualitative results for datasets, respectively in Fig. 5 and Fig. 6 to further support the effectiveness and superiority of the proposed MegAD model.

Table 9. Comparison of image-level anomaly detection in terms of AUROC on MVTec-AD.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
bottle	94.2	92.9	99.4	99.2	99.4	99.5	99.8	96.8	99.2	99.6
cable	49.7	46.1	88.8	60.8	58.8	89.5	94.2	74.3	87.6	91.4
capsule	49.9	46	67.8	68.5	57.5	65.4	84.6	52.5	90.7	92.5
carpet	99.6	99.2	95.3	95.0	93.2	100.0	100.0	100.0	99.9	99.9
grid	58.1	48.0	63.6	73.6	56.4	99.8	99.8	100.0	100.0	99.4
hazelnut	94.5	93.7	88.3	97.5	90.0	99.5	99.8	78.3	100.0	100.0
leather	98.3	97.7	97.3	98.1	99.1	100.0	100.0	100.0	100.0	100.0
metal_nut	48.7	50.5	73.4	83.2	49.0	100.0	99.1	96.3	100.0	100.0
pill	56.8	54.5	81.9	64.4	81.7	95.3	92.6	82.5	94.5	94.5
screw	56.2	51.5	44.4	57.0	58.8	77.4	65.0	37.8	66.9	81.8
tile	78.4	77.7	99.0	94.7	91.3	99.8	100.0	99.4	99.8	99.8
toothbrush	73.9	73.1	83.3	79.3	69.7	97.8	98.9	87.2	96.4	98.9
transistor	77.2	76.6	78.1	67.4	66.6	91.8	94.0	52.3	91.0	94.5
wood	97.5	98.0	97.8	99.3	96.4	99.7	97.9	99.3	98.4	98.8
zipper	64.5	62.5	92.3	80.7	88.8	97.9	93.9	99.0	98.7	99.6
mean	73.2	71.2	83.4	81.2	77.1	94.2	94.6	83.7	94.8	96.7

Table 10. Comparison of image-level anomaly detection in terms of Accuracy on MVTec-AD.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
bottle	53.1	34.9	91.2	93.5	88.9	98.8	100.0	99.1	97.6	98.8
cable	64.6	43.3	58.4	61.9	76.9	78.7	94.8	83.6	80.7	87.3
capsule	36.8	18.9	69.4	72.4	96.8	74.2	100.0	86.7	83.3	85.6
carpet	54.2	34.2	71.6	74.7	86.3	99.2	99.1	100.0	99.1	99.2
grid	47.6	29.5	70.8	73.5	92.4	92.3	96.2	100.0	100.0	98.7
hazelnut	62.9	46.4	84.6	87.8	97.6	78.2	98.8	88.6	100.0	99.1
leather	56.3	36.3	50.6	53.5	62.7	98.4	84.0	100.0	100.0	100.0
metal_nut	39.2	20.9	67.0	70.4	82.6	93.9	81.8	99.2	100.0	100.0
pill	39.7	23.4	28.5	31.5	82.7	88.0	98.2	96.3	88.0	85.0
screw	52.9	33.1	70.9	74.7	81.7	69.4	90.4	65.4	68.75	70.0
tile	56.1	38.5	53.8	56.8	86.8	96.6	81.4	99.7	99.1	99.2
toothbrush	56.1	40.5	71.7	74.0	75.0	81.0	68.7	95.1	92.9	97.6
transistor	82.4	66.0	47.5	48.7	71.4	81.0	85.7	43.9	85.0	89.0
wood	54.2	34.2	86.6	91.6	56.0	88.6	79.0	89.7	95.0	96.2
zipper	48.0	30.5	58.2	60.7	89.4	77.5	75.5	87.3	90.7	98.7
mean	53.6	35.4	65.4	68.4	81.8	86.4	88.9	89.0	92.0	93.6

Table 11. Comparison of pixel-level anomaly detection in terms of AUROC on MVTec-AD.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
bottle	60.8	80.1	97.9	98.4	97.7	98.0	99.6	95.5	97.9	98.9
cable	57.1	78.6	95.5	92.9	76.6	93.6	98.4	82.7	91.8	93.2
capsule	70.8	88.7	95.6	96.5	98.5	89.8	99.5	94.3	97.0	98.5
carpet	78.5	98.3	98.4	98.7	82.9	99.4	95.9	99.5	99.4	100.2
grid	36.5	49.4	58.8	69.9	92.2	97.3	95.1	97.3	97.2	98.2
hazelnut	71.7	94.2	95.8	97.9	90.9	98.9	97.3	95.8	98.8	99.7
leather	62.2	79.7	98.8	99.3	55.4	99.3	93.3	99.3	99.3	100.0
metal_nut	45.3	66.6	89.3	95.8	89.8	90.4	97.3	79.0	92.5	95.3
pill	58.9	78.9	93.1	96.7	98.0	96.0	98.4	86.8	97.0	97.5
screw	66.5	83.9	89.6	93.5	65.3	96.3	91.4	92.4	96.0	98.3
tile	61.2	84.0	94.1	94.1	93.6	95.9	92.8	95.1	96.6	98.5
toothbrush	68.0	92.5	97.3	97.6	93.5	98.4	94.0	96.5	98.9	99.5
transistor	65.3	85.3	84.9	83.5	94.9	87.9	99.1	73.4	83.0	84.4
wood	73.1	91.7	92.7	95.8	68.5	96.0	89.4	95.6	95.8	97.2
zipper	68.8	90.3	97.4	94.5	95.7	94.7	96.6	97.6	94.7	98.6
mean	62.9	84.0	89.9	93.7	86.2	95.4	95.9	92.1	95.7	97.2

Table 12. Comparison of image-level anomaly detection in terms of AUROC on MVTec-AD2.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
can	23.4	42.8	34.5	35.7	28.2	31.0	31.8	23.8	39.0	42.1
fabric	38.4	36.7	49.5	50.7	58.4	51.9	51.3	42.7	51.0	51.5
fruit_jelly	59.9	38.8	71.0	72.2	75.3	75.4	86.9	67.3	73.8	76.7
rice	47.2	30.8	58.3	59.5	46.3	64.6	73.2	67.7	60.4	62.0
sheet_metal	62.5	81.3	73.6	74.8	55.3	84.4	60.4	58.0	86.3	86.5
vial	62.7	47.9	73.8	75.0	73.4	79.2	75.3	62.4	87.4	88.2
wallplugs	38.0	50.6	49.1	50.3	62.6	37.9	48.6	49.3	45.4	46.2
walnuts	60.6	66.0	71.7	72.9	69.9	78.9	72.8	50.8	80.2	80.5
mean	49.1	49.4	60.2	61.4	58.7	62.9	62.5	52.8	65.4	66.6

Table 13. Comparison of image-level anomaly detection in terms of Accuracy on MVTec-AD2.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
can	33.4	43.8	39.8	42.4	56.2	49.4	56.2	40.6	45.7	52.0
fabric	37.7	39.7	43.5	48.0	60.9	46.8	50.0	52.6	47.4	59.6
fruit_jelly	35.9	25.0	34.2	55.3	78.8	73.8	78.8	82.5	66.2	73.3
rice	32.6	33.3	37.9	52.1	68.2	32.6	67.4	72.3	54.6	66.1
sheet_metal	35.2	31.6	49.1	59.5	79.0	50.0	57.9	76.6	72.0	79.8
vial	45.5	32.1	52.4	64.6	75.0	68.6	80.0	82.8	82.9	89.1
wallplugs	46.3	50.0	55.6	45.9	65.3	34.0	52.7	55.6	48.7	57.0
walnuts	41.0	43.3	47.5	54.7	65.3	70.7	73.3	60.8	72.0	79.0
mean	38.5	37.4	45.0	52.8	68.6	53.2	64.5	65.5	61.1	69.5

Table 14. Comparison of pixel-level anomaly detection in terms of AUROC on MVTec-AD2.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
can	61.5	66.8	76.2	77.2	80.7	74.4	77.5	85.9	60.8	69.4
fabric	62.9	67.3	74.6	78.5	84.1	76.3	74.6	69.1	78.2	86.3
fruit_jelly	68.3	88.9	81.7	83.3	90.6	89.3	90.5	84.3	86.7	95.5
rice	59.2	76.9	81.1	84.1	64.2	66.9	90.0	67.8	63.1	72.3
sheet_metal	56.8	64.4	72.9	75.6	74.5	50.4	84.3	92.5	51.7	60.4
vial	69.6	85.6	84.5	87.0	84.9	89.9	92.5	87.9	90.5	99.2
wallplugs	65.1	75.5	77.3	77.9	73.1	78.1	75.4	68.7	73.0	81.8
walnuts	70.9	89.3	80.0	81.8	82.5	89.4	91.9	88.7	89.8	98.6
mean	64.3	76.8	78.5	80.7	79.3	76.8	83.2	80.6	74.2	82.9

Table 15. Comparison of image-level anomaly detection in terms of AUROC on VisA.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
Candle	66.2	25.2	85.1	72.8	84.2	93.5	88.7	68.6	96.0	97.4
Capsules	69.5	67.0	60.0	58.8	58.7	90.6	75.0	50.1	88.4	90.0
Cashew	70.0	55.0	89.5	79.0	66.4	92.0	92.7	62.0	94.6	95.6
Chewinggum	72.1	48.0	95.3	87.8	76.2	98.6	96.1	86.7	98.7	100.0
Fryum	67.3	58.8	77.0	73.0	67.8	94.8	92.1	71.6	93.6	95.3
Macaronil	67.9	67.4	68.0	63.2	57.7	92.1	84.1	74.0	96.6	97.6
Macaroni2	63.8	51.5	55.6	51.9	56.6	79.5	79.2	45.9	84.2	87.4
Pcb1	76.4	63.5	78.9	74.4	76.6	84.0	89.9	80.6	84.5	81.4
Pcb2	73.2	56.0	81.5	64.8	58.6	68.7	74.1	55.6	83.4	81.1
Pcb3	74.6	57.8	82.7	60.2	52.3	79.7	80.4	55.5	81.6	82.7
Pcb4	78.0	61.9	93.9	77.6	80.3	76.5	92.9	49.6	97.9	98.9
Pipe_fryum	75.5	56.5	90.7	81.7	92.0	99.3	96.5	96.9	99.5	100.0
mean	71.2	55.7	74.8	70.4	69.0	87.4	86.8	66.4	91.6	92.4

Table 16. Comparison of image-level anomaly detection in terms of Accuracy on VisA.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
Candle	47.3	42.0	70.1	65.3	74.5	86.5	82.5	68.9	91.0	90.7
Capsules	50.7	45.0	67.4	62.5	62.5	75.0	70.6	63.5	81.9	84.3
Cashew	47.3	42.0	74.8	69.4	68.7	58.0	86.7	79.9	87.3	89.2
Chewinggum	42.8	38.0	85.1	79.0	72.7	96.0	91.3	93.9	93.3	97.2
Fryum	44.3	39.3	75.7	70.3	70.0	79.3	88.7	85.1	90.7	91.9
Macaronil	59.7	53.0	55.2	51.2	52.0	63.0	76.5	76.5	90.0	91.2
Macaroni2	56.3	50.0	53.9	50.0	57.5	59.0	72.0	51.0	79.0	83.2
Pcb1	57.4	51.0	55.2	51.2	74.0	76.0	81.5	74.8	81.5	82.7
Pcb2	59.7	53.0	55.5	51.5	55.5	58.5	70.5	57.9	78.5	75.7
Pcb3	57.1	50.7	55.1	51.1	53.7	72.1	76.1	53.0	73.1	74.8
Pcb4	60.5	53.7	62.9	58.4	71.6	79.1	87.1	55.0	94.5	95.7
Pipe_fryum	48.8	43.3	70.2	65.2	86.7	92.7	88.7	98.4	98.7	99.9
mean	52.7	46.8	65.2	60.4	66.6	74.6	81.0	71.5	86.6	88.0

Table 17. Comparison of pixel-level anomaly detection in terms of AUROC on VisA.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
Candle	62.2	80.9	96.9	94.5	90.8	98.5	62.0	97.2	98.6	99.5
Capsules	56.9	67.2	92.2	86.3	82.9	97.3	96.8	91.3	97.5	98.3
Cashew	67.4	91.8	98.1	96.2	91.8	96.9	95.9	90.8	96.9	96.4
Chewinggum	54.5	89.5	95.9	97.8	97.0	99.0	96.3	99.1	99.2	100.0
Fryum	66.1	86.2	93.3	95.1	90.8	93.6	95.2	86.5	93.8	94.7
Macaronil	70.3	86.9	95.2	90.9	91.4	97.5	95.4	91.4	98.6	99.1
Macaroni2	65.0	85.0	89.1	93.4	84.0	95.4	95.7	90.3	97.3	98.2
Pcb1	68.5	80.4	96.1	95.8	84.6	97.6	90.5	75.8	97.3	98.8
Pcb2	70.8	84.6	95.4	91.7	85.6	93.9	99.3	81.2	94.0	93.9
Pcb3	72.2	87.7	96.2	93.7	86.9	94.3	98.2	86.0	95.6	97.0
Pcb4	60.7	76.2	95.6	88.5	94.3	94.4	92.5	91.4	96.0	95.7
Pipe_fryum	69.2	95.9	98.8	99.1	98.1	98.2	62.0	95.4	98.3	99.6
mean	65.3	84.4	93.4	93.6	89.8	96.4	96.8	89.7	97.0	97.6

Table 18. Comparison of image-level anomaly detection in terms of AUROC on DAGM.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
Class1	47.4	44.1	63.1	68.9	57.3	65.9	88.9	76.4	80.8	92.8
Class2	84.3	95.2	89.3	97.6	77.2	100.0	99.7	72.1	100.0	100.0
Class3	43.0	55.7	74.4	94.1	78.7	100.0	100.0	98.9	100.0	100.0
Class4	85.8	89.3	88.6	99.4	58.4	100.0	99.9	47.9	100.0	100.0
Class5	50.1	61.5	65.8	66.9	69.9	99.9	98.9	99.0	99.9	99.9
Class6	73.2	81.8	79.7	91.2	70.9	99.6	100.0	71.4	99.9	99.9
Class7	42.6	51.1	77.2	88.5	56.3	100.0	100.0	99.9	100.0	100.0
Class8	45.8	53.6	56.0	53.3	51.5	89.2	83.3	59.8	91.1	91.6
Class9	87.3	95.7	87.5	98.8	64.9	99.9	98.1	98.6	99.9	99.9
Class10	68.5	72.1	75.1	84.9	74.9	100.0	100.0	62.6	100.0	100.0
mean	62.8	70.0	75.7	84.4	66.0	95.5	96.9	78.7	97.1	98.4

Table 19. Comparison of image-level anomaly detection in terms of Accuracy on DAGM.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
Class1	44.4	49.3	66.2	63.9	53.3	55.4	87.0	73.9	82.6	90.0
Class2	55.1	60.0	92.8	90.5	69.3	84.7	98.7	71.1	100.0	100.0
Class3	46.4	51.3	77.4	75.1	72.7	98.7	99.3	99.1	100.0	100.0
Class4	55.1	60.0	84.2	81.9	50.0	100.0	98.7	46.5	100.0	100.0
Class5	79.7	84.6	62.4	60.1	56.3	94.0	97.0	97.2	98.3	99.0
Class6	51.8	56.7	63.3	61	64.3	98.0	100.0	73.2	99.0	99.3
Class7	45.8	50.7	77.0	74.7	54.3	98.7	100.0	99.9	100.0	100.0
Class8	46.1	51.0	52.5	50.2	50.5	76.2	79.7	58.9	87.3	87.0
Class9	55.6	60.0	94.5	92.2	54.3	95.0	93.0	98.8	99.0	98.5
Class10	54.5	59.0	67.8	65.5	65.5	98.8	99.3	63.7	99.5	99.8
mean	53.4	55.7	73.8	71.5	59.1	89.9	95.3	78.2	96.6	97.4

Table 20. Comparison of pixel-level anomaly detection in terms of AUROC on DAGM.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
Class1	50.6	62.1	70.4	79.6	80.0	62.0	83.9	83.7	70.4	73.0
Class2	78.0	96.1	88.7	98.5	87.9	96.8	98.5	96.5	90.3	96.8
Class3	64.2	75.7	83.0	92.7	92.2	95.9	97.9	96.8	95.6	96.3
Class4	78.8	96.4	86.8	99.4	90.3	96.3	98.4	88.7	96.0	97.3
Class5	71.3	58.7	82.1	82.8	91.3	95.2	96.0	95.6	95.5	96.4
Class6	69.1	87.0	86.3	90.3	80.6	95.4	97.0	94.9	95.5	96.5
Class7	73.3	68.6	86.7	86.9	84.8	95.7	95.8	95.7	95.9	96.4
Class8	57.9	64.3	80.6	67.4	69.4	90.5	92.5	89.8	90.5	91.5
Class9	78.4	99.4	90.3	99.9	89.9	99.3	99.8	99.5	99.4	99.4
Class10	79.6	89.6	87.6	97.1	91.1	98.2	98.6	96.8	98.1	98.5
mean	70.1	82.4	84.3	89.4	85.7	92.5	94.9	93.8	93.3	94.2

Table 21. Comparison of image-level anomaly detection in terms of AUROC on MPDD.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
bracket_black	41.5	45.1	65.8	59.3	47.9	43.5	69.0	40.7	57.9	61.8
bracket_brown	49.7	51.7	57.5	51.0	53.2	58.4	71.4	49.8	47.4	64.1
bracket_white	67.1	69.1	50.3	43.8	37.2	47.8	56.6	73.6	61.0	60.7
connector	19.2	21.2	60.9	54.4	65.0	92.9	84.5	61.2	73.8	79.0
metal_plate	30.6	32.6	61.3	54.8	81.2	100.0	94.8	94.5	99.9	100.0
tubes	56.3	58.3	72.8	66.3	78.7	93.3	85.4	96.0	98.3	95.5
mean	44.0	46.3	61.4	54.9	60.5	72.6	76.9	69.3	73.0	76.9

Table 22. Comparison of image-level anomaly detection in terms of Accuracy on MPDD.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
bracket_black	40.7	40.5	64.4	60.1	64.6	60.8	64.6	54.2	62.0	67.1
bracket_brown	36.6	36.4	67.2	66.5	67.5	66.2	75.3	71.1	70.1	78.0
bracket_white	53.5	53.3	49.8	50.0	50.0	50.0	68.3	75.9	65.0	66.7
connector	57.0	56.8	63.2	31.8	63.6	79.6	79.6	41.9	63.6	79.6
metal_plate	27.0	26.8	80.3	73.2	81.4	87.6	92.8	97.9	99.0	100.0
tubes	42.8	42.6	76.9	68.3	77.2	91.1	77.2	88.3	95.0	88.1
mean	42.9	42.7	67.0	58.3	67.4	72.5	76.3	71.6	75.8	78.9

Table 23. Comparison of pixel-level anomaly detection in terms of AUROC on MPDD.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
bracket_black	82.5	86.8	90.8	92.8	87.6	90.2	94.1	91.1	93.6	93.4
bracket_brown	81.0	91.4	92.1	94.0	93.7	94.4	94.1	87.4	94.7	96.0
bracket_white	78.2	94.4	88.8	90.2	87.9	97.0	96.4	98.1	97.5	98.4
connector	80.8	91.3	92.5	95.0	90.4	98.0	98.2	95.3	97.5	97.0
metal_plate	75.5	80.3	88.0	90.7	92.8	98.0	95.3	96.4	98.1	98.2
tubes	80.2	84.0	89.4	92.7	95.7	99.0	98.2	99.0	99.5	99.0
mean	79.7	88.0	90.3	92.6	91.3	96.1	93.3	94.6	96.8	97.0

Table 24. Comparison of image-level anomaly detection in terms of AUROC on MulSen-AD.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
botton_cell	52.4	75.0	60.0	66.0	60.0	100.0	86.2	70.7	100.0	100.0
capsule	49.9	66.1	63.6	79.3	25.1	85.6	99.5	72.2	86.3	84.9
cotton	56.8	61.3	76.8	79.8	68.8	99.4	100.0	100.0	100.0	100.0
cube	31.7	33.0	79.9	79.8	87.3	92.7	100.0	95.8	100.0	96.7
flat_pad	65.0	92.3	66.5	67.5	55.0	83.0	87.7	78.3	89.3	90.3
light	60.6	77.5	74.1	52.3	59.2	96.7	72.5	48.3	87.5	96.7
nut	46.5	51.7	55.3	34.6	100.0	68.5	54.8	79.3	67.2	93.4
piggy	54.3	60.3	81.2	72.2	51.4	97.6	92.4	64.8	97.2	97.0
plastic_cylinder	68.7	91.2	85.7	73.3	68.8	100.0	93.5	92.9	95.9	100.0
screen	51.0	54.5	52.7	48.8	64.5	68.8	69.0	58.6	93.1	73.5
screw	55.2	62.3	20.0	78.8	81.6	23.9	99.0	52.3	10.3	96.8
solar_panel	61.9	69.2	72.8	44.0	68.3	95.0	64.2	96.7	85.0	72.5
spring_pad	31.5	0.4	67.3	57.3	34.6	88.5	77.5	40.4	85.0	87.5
toothbrush	60.1	67.5	50.5	49.8	41.0	64.0	70.0	59.8	81.0	59.5
zipper	50.7	56.8	69.9	65.1	77.9	81.6	85.3	86.8	83.2	99.5
mean	53.1	61.3	65.1	63.2	62.9	83.0	83.4	71.2	84.0	89.9

Table 25. Comparison of image-level anomaly detection in terms of Accuracy on MulSen-AD.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
botton_cell	52.1	38.9	64.8	69.4	72.2	55.6	77.8	80.7	100.0	100.0
capsule	45.5	27.5	71.5	76.9	80.4	86.3	98.0	92.4	76.5	78.4
cotton	57.0	55.6	66.3	82.6	66.7	61.1	100.0	100.0	100.0	100.0
cube	32.2	20.9	78.6	82.8	93.0	79.1	100.0	98.8	100.0	93.0
flat_pad	47.0	35.0	71.8	74.9	80.0	67.5	75.0	90.4	90.0	92.5
light	61.8	59.1	61.3	64.7	54.6	63.6	77.3	53.1	81.8	90.9
nut	36.5	30.8	66.6	69.2	100.0	51.3	48.7	84.5	49.0	87.2
piggy	42.4	35.9	76.5	72.5	76.9	84.6	84.6	87.4	94.9	89.7
plastic_cylinder	55.1	48.1	81.1	80.4	70.4	100.0	92.6	97.1	96.3	100.0
screen	49.7	35.9	68.6	70.7	74.4	94.9	66.7	86.2	92.3	76.9
screw	47.0	36.6	74.3	82.0	82.9	73.2	97.6	75.8	26.8	97.6
solar_panel	61.7	50.0	64.7	61.8	59.1	63.6	68.2	77.4	86.4	72.7
spring_pad	27.5	17.6	61.4	66.3	70.6	52.9	82.4	71.8	79.4	79.4
toothbrush	52.3	43.3	66.9	66.1	66.7	63.3	76.7	81.6	80.0	56.7
zipper	57.0	44.8	72.2	76.4	69.0	79.3	82.8	92.7	86.2	96.6
mean	48.2	38.7	69.7	73.0	74.5	71.8	81.9	84.0	83.3	87.4

Table 26. Comparison of pixel-level anomaly detection in terms of AUROC on MulSen-AD.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
botton_cell	88.9	99.4	99.0	97.3	99.0	99.1	99.3	99.2	99.4	99.2
capsule	85.6	97.8	97.4	97.3	97.4	99.1	99.1	98.1	99.4	98.6
cotton	82.2	94.6	94.4	97.1	94.4	98.9	99.3	60.1	99.0	98.9
cube	78.8	91.2	80.8	96.2	80.8	98.0	97.8	99.3	97.7	97.7
flat_pad	84.5	95.9	97.9	95.6	97.9	97.4	98.3	98.5	98.1	97.8
light	81.1	92.4	92.0	93.7	92.0	95.5	96.8	94.2	95.2	95.5
nut	85.4	97.4	98.4	96.3	98.4	98.1	99.1	98.3	98.7	98.2
piggy	72.7	82.4	92.4	93.4	92.4	95.2	96.2	95.1	95.0	94.3
plastic_cylinder	80.9	92.0	97.3	94.7	97.3	96.5	96.2	98.5	97.7	97.1
screen	67.5	76.7	74.4	86.8	74.4	88.6	82.7	93.8	87.5	87.2
screw	86.3	98.1	98.5	94.9	98.5	96.7	99.0	98.5	97.3	97.4
solar_panel	77.6	88.8	92.3	93.0	92.3	94.8	96.8	94.3	90.0	89.6
spring_pad	85.4	97.4	98.4	97.3	98.4	99.1	99.1	98.7	99.1	99.1
toothbrush	85.0	96.8	97.7	95.7	97.7	97.5	98.4	98.8	96.4	94.9
zipper	76.5	87.7	95.7	86.5	95.7	88.3	94.7	93.8	92.0	88.5
mean	81.2	92.6	93.9	94.4	93.8	96.2	96.4	97.2	96.1	95.6

Table 27. Comparison of image-level anomaly detection in terms of AUROC on medical.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
BrainMRI	49.4	61.2	73.2	44.0	51.7	73.1	70.5	52.5	72.2	74.0
LiverCT	41.0	57.1	44.9	58.4	62.8	60.3	61.7	63.6	61.4	63.4
RESC	52.8	53.0	56.3	68.7	63.9	82.4	85.5	63.5	85.7	86.3
mean	47.7	57.1	58.1	57.0	59.5	71.9	72.5	59.9	73.1	74.6

Table 28. Comparison of image-level anomaly detection in terms of Accuracy on medical.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
BrainMRI	73.8	75.7	84.0	75.4	82.9	90.1	76.3	88.4	88.6	90.9
LiverCT	40.5	51.2	48.7	53.9	56.4	63.8	61.2	48.8	64.2	64.3
RESC	49.3	45.6	55.1	60.2	52.1	62.3	79.5	78.3	65.8	78.4
mean	54.5	57.5	62.6	63.2	63.8	72.1	72.3	71.8	72.9	77.9

Table 29. Comparison of pixel-level anomaly detection in terms of AUROC on medical.

Category	SPADE	PaDiM	PatchCore	RegAD	WinCLIP	AnomalyGPT	PromptAD	ResAD†	KAG-prompt	**Ours**
BrainMRI	81.2	88.4	96.0	83.1	84.9	96.0	94.6	91.9	95.3	96.5
LiverCT	89.6	86.0	95.6	91.3	98.2	95.8	96.7	96.9	96.7	98.3
RESC	71.9	72.5	78.2	85.6	90.3	94.0	96.0	81.7	96.1	96.8
mean	80.9	82.3	89.9	86.7	91.1	95.3	95.7	90.2	96.0	97.2

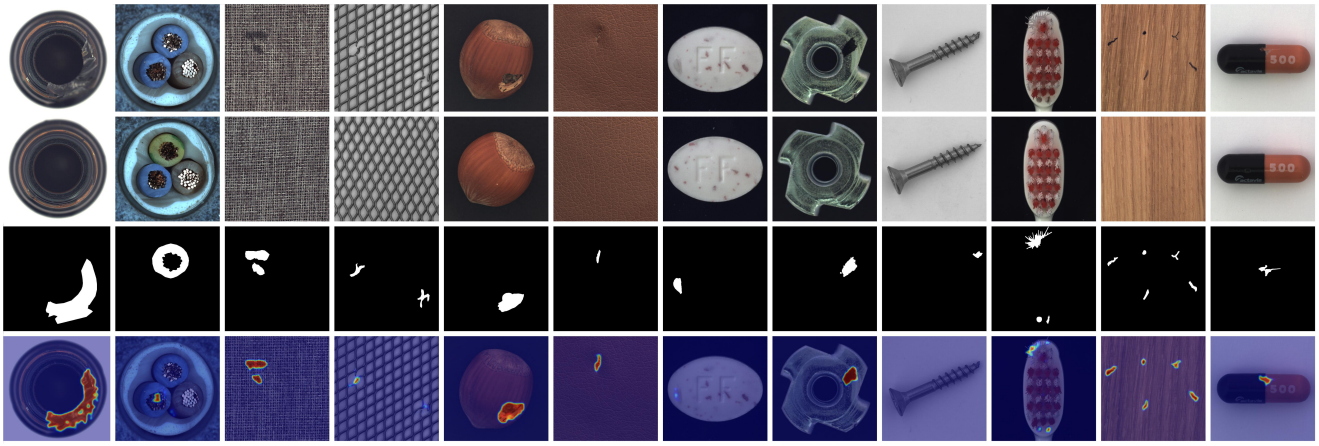


Figure 5. Additional qualitative results on MVTec-AD dataset, From top to bottom: the original input image (with anomalies), the normal image, the ground truth mask, and the anomaly heatmaps.

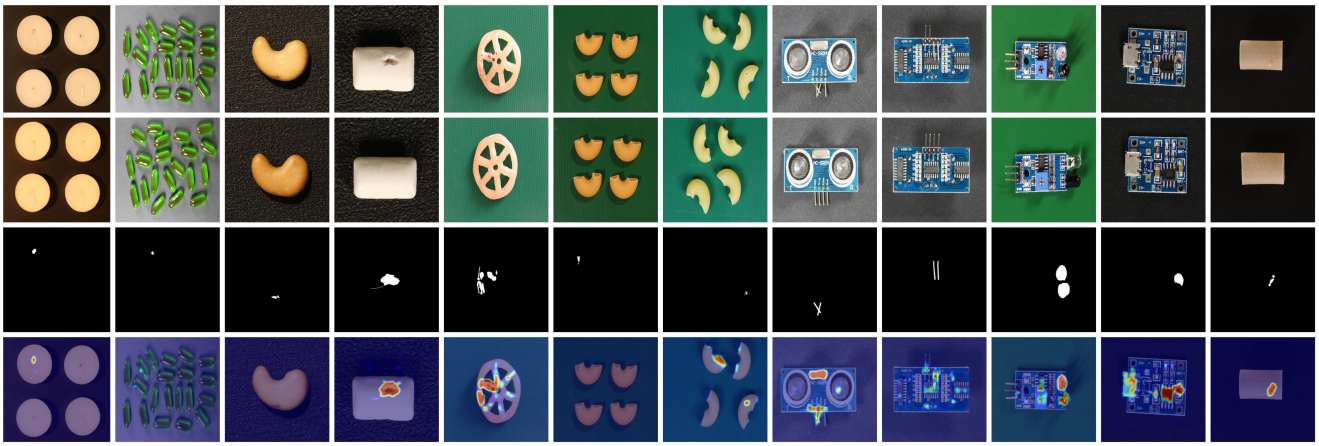


Figure 6. Additional qualitative results on VisA dataset, From top to bottom: the original input image (with anomalies), the normal image, the ground truth mask, and the anomaly heatmaps.