Uncertainty-Guided Dual-View Distillation for Industrial Anomaly Detection

Conghua Wei[1], Xidong Xi[1], Guitao Cao[1*], Mengshang Nie[1]
[1]Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, China
Email: {51275902175, 52265902004}@stu.ecnu.edu.cn, gtcao@sei.ecnu.edu.cn, 51275902051@stu.ecnu.edu.cn

***Abstract*—Teacher-Student (T-S) architectures are promising for industrial anomaly detection but often suffer from "over-generalization", where students learn trivial local identity mappings that bypass global spectral constraints, allowing them to undesirably replicate high-frequency anomalies. To address this issue, we propose a novel Dual-View Distillation Framework (DVDF). First, we introduce an Omni-Frequency Student (OFS) that enforces consistency across spatial and frequency domains, effectively inhibiting the replication of anomalous patterns. Second, we employ an Uncertainty-Guided Distillation (UGD) strategy via Monte Carlo Dropout to suppress unreliable supervision signals, preventing the student from mimicking noisy teacher predictions. Finally, we design a Cross-scale Efficiency Block (CEB) to integrate multi-scale features, preserving salient fine-grained details often lost in standard bottlenecks. Results on HSS-IAD and MVTec AD benchmarks indicate that our method achieves highly competitive performance compared to current state-of-the-art methods.**

***Index Terms*—Anomaly detection, Knowledge distillation, Teacher-student model, Dual-View Network, Uncertainty**

I. Introduction

Industrial anomaly detection aims to identify and locate rare defects, which is a critical task for automated quality control. While feature embedding methods using memory banks or normalizing flows have set high benchmarks, they frequently suffer from prohibitive memory consumption or latency. Consequently, Teacher-Student (T-S) architectures have emerged as a dominant paradigm. Unlike memory-heavy approaches that require extensive retrieval during inference, T-S models encapsulate normality within network weights. By distinguishing anomalies through knowledge distillation discrepancies, T-S models offer a compelling balance between accuracy and efficiency, making them highly suitable for industrial deployment.

The T-S architecture achieves anomaly detection via knowledge distillation, creating feature-level discrepancies to identify defects. Typically, the teacher is a pretrained deep model with robust feature extraction capabilities, while the student is a lightweight network trained to mimic the teacher's representations on normal samples. Essentially, the anomaly detection task is reformulated as a feature regression problem where the student strives to generalize. Since the student never encounters anomalies during training, it theoretically fails to replicate the teacher's features for anomalous inputs, yielding detectable discrepancies at inference time. Representative works include STPM [1], RD [2], and RD++ [3]. A schematic comparison between these standard paradigms and our proposed framework is illustrated in Fig. 1.

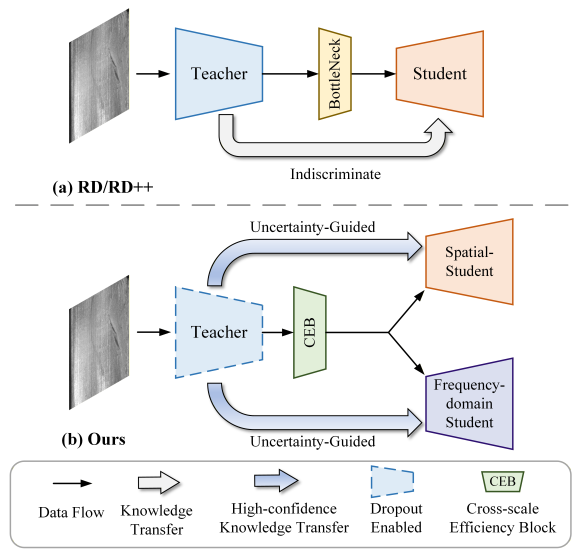

Fig. 1. Comparison between **(a)** standard Reverse Distillation (RD/RD++) and **(b)** our proposed framework. Unlike the single-stream student in RD, our method introduces a dual-view student architecture and leverages uncertainty-guided distillation to enforce high-confidence knowledge transfer.

However, strictly enforcing normality in T-S architectures requires addressing three fundamental constraints. First, preventing the student from learning trivial "identity shortcuts" is paramount. Since standard CNNs with local receptive fields may inadvertently replicate high-frequency anomalies (e.g., scratches) rather than suppressing them, it is essential to impose constraints beyond the spatial domain—specifically, by leveraging global frequency consistency to distinguish anomalies. Second, preserving fine-grained details is critical for industrial defects. Rigid bottlenecks in prior arts often aggressively compress features, causing subtle defects to vanish; thus, a robust architecture must dynamically synthesize multi-scale information rather than discarding it. Third, distillation must account for the reliability of supervision. Given the domain shift between natural and industrial images, the teacher's features inevitably contain noise. Instead of treating all supervision as equal, an ideal strategy should quantify uncertainty to prevent the student from overfitting to unreliable feature regions.

Motivated by these observations, we propose the Dual-View

*Corresponding author. Email: gtcao@sei.ecnu.edu.cn

Distillation Framework (DVDF), trained via an Uncertainty-Guided Distillation (UGD) strategy. Structurally, the DVDF integrates two key components to construct a strict knowledge bottleneck: first, an Omni-Frequency Student (OFS) that enforces consistency across spatial and frequency domains to prevent the "identity shortcut" common in pure CNNs; second, a Cross-scale Efficiency Block (CEB) that refines multi-scale teacher features to preserve salient details by establishing a dense, information-rich input representation. Strategically, drawing on Bayesian uncertainty modeling [4], our UGD strategy treats the teacher's predictions as probabilistic. By assigning lower weights to noisy regions based on inverse variance, we prevent the student from overfitting to unreliable supervision.

Our contributions are summarized as follows:

- We propose the **Omni-Frequency Student (OFS)** to overcome CNN spectral bias, utilizing dynamic spatial-frequency gating to strictly inhibit the replication of anomalies.
- We devise an **Uncertainty-Guided Distillation (UGD)** strategy that suppresses unreliable supervision signals by quantifying teacher uncertainty, preventing overfitting to noisy regions.
- We introduce the **Cross-scale Efficiency Block (CEB)** to address bottleneck information loss, integrating multi-scale teacher features to preserve salient fine-grained details.

Our method demonstrates strong performance on the MVTec AD [5] and HSS-IAD [6] benchmark datasets. It excels especially on the HSS-IAD dataset, which is dominated by metal casting defects, demonstrating its capability in identifying small and easily confusable anomalies. These results validate the synergistic effectiveness of our dual-view design, multi-scale feature integration, and uncertainty-guided distillation strategy in handling complex real-world industrial scenarios.

II. Related Work

A. Unsupervised Industrial Anomaly Detection

Due to the scarcity of abnormal samples, the mainstream approaches typically employ unsupervised anomaly detection methods that only use normal samples for training [7], [8]. Existing approaches can be broadly categorized into image reconstruction and feature embedding methods.

Reconstruction-based methods assume that models trained on normal data can reconstruct normal regions well but fail on anomalies. Early works employed Autoencoders (AE), Generative Adversarial Networks (GANs), and more recently, Transformers and Diffusion Models to restore input images [9]–[11]. Most reconstruction methods are trained from scratch without leveraging powerful pretrained models, resulting in comparatively poorer performance relative to image-level feature embedding techniques.

Feature embedding methods leverage pretrained networks to extract discriminative representations. One-class classification approaches map features into a compact sphere [12]. Distribution-based methods employ Normalizing Flows to estimate the likelihood of feature density [13]. Memory bank methods [14] store a large database of normal feature patches for nearest-neighbor matching. While these methods achieve state-of-the-art performance, their high memory consumption and inference latency hinder deployment in resource-constrained industrial environments.

B. Knowledge Distillation for Anomaly Detection

This framework transfers knowledge from a fixed pre-trained teacher to a lightweight student trained only on normal samples, eliminating the high memory overhead of memory bank approaches. During inference, anomalies are identified by measuring the discrepancy between teacher and student outputs—normal regions yield similar responses, while anomalous regions produce significant deviations.

Pioneering works like STPM [1] utilize multi-scale feature alignment to capture defects of various sizes across different network hierarchies. More recent advances, such as RD [2] and RD++ [3], introduce heterogeneous architectures with reversed information flow to prevent the student from merely copying the teacher. By designing asymmetric structures, these methods ensure the student learns meaningful representations of normality rather than memorizing outputs, thereby improving discriminative power for anomaly detection.

III. Methodology

As illustrated in Figure 2, our approach is established upon two parts: the Dual-View Distillation Framework (DVDF) and the Uncertainty-Guided Distillation (UGD) strategy.

A. Dual-View Distillation Framework (DVDF)

The DVDF is orchestrated through three core components designed to enforce strict normality constraints: the Cross-scale Efficiency Block (CEB) for multi-scale feature refinement, the Omni-Frequency Student (OFS) for dual-domain learning, and the Dynamic Gated Fusion (DGF) for adaptive integration. The details of each module are described as follows.

1) Cross-scale Efficiency Block (CEB): The CEB is designed to intelligently refine the multi-scale features of the teacher model into a dense, information-rich representation F_{in}. This representation then serves as the input for the Omni-Frequency Student. The CEB begins by unifying the channel dimensions and spatial resolutions of the multi-scale features. These aligned features are then processed through iterative top-down and bottom-up pathways to ensure a rich exchange between semantic and spatial information. Finally, the fused features are concatenated and fed into a compression module, which uses efficient depthwise separable convolutions and Squeeze-and-Excitation (SE) [15] attention blocks to recalibrate channel importance, ultimately producing the dense bottleneck representation F_{in}.

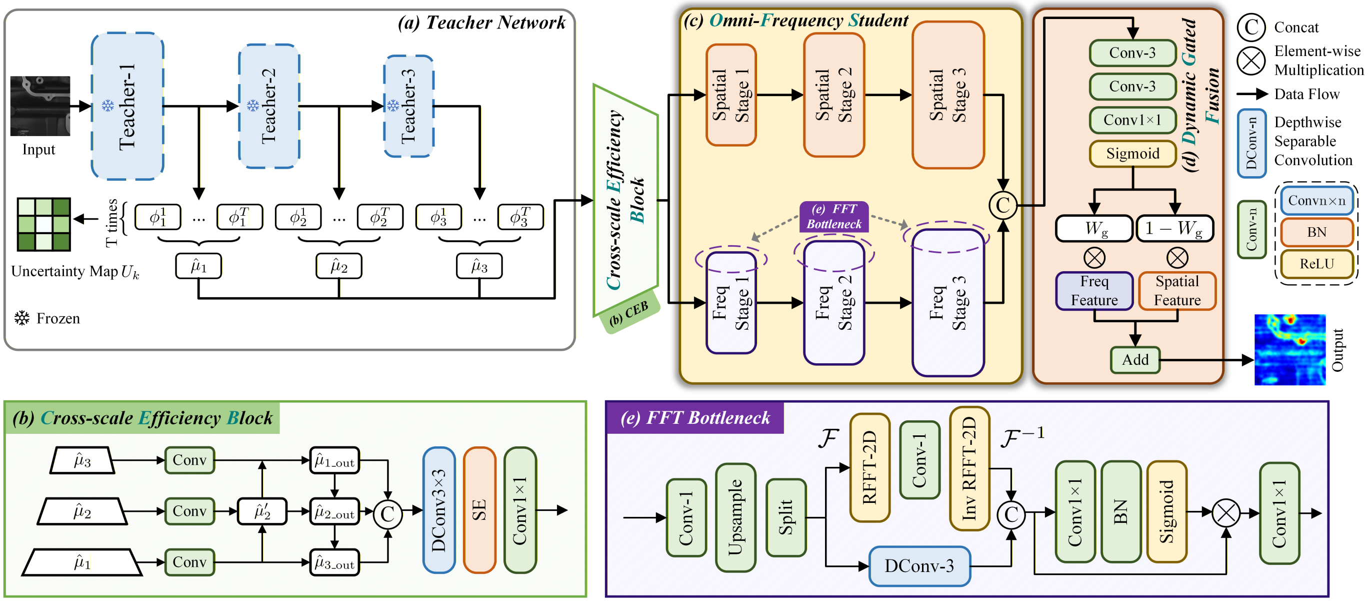

Fig. 2. Schematic diagram of the DVDF framework, consisting of: **(a)** the uncertainty-guided teacher network; **(b)** the Cross-scale Efficiency Block (CEB); **(c)** the Omni-Frequency Student (OFS); **(d)** the Dynamic Gated Fusion (DGF) module; and **(e)** the FFT Bottleneck.

2) Omni-Frequency Student (OFS): Traditional knowledge distillation configures the student model as a single decoder, oversimplifying the learning task to merely approximating the feature embedding capability of the teacher model. To address this issue, we propose the OFS architecture with spatial and frequency branches. The spatial branch D_{spatial} captures fine-grained local details, while the frequency branch D_{freq} models global structural composition from the input F_{in}. Their respective outputs are defined as:

$$F_{\text{spatial}} = D_{\text{spatial}}(F_{\text{in}}), \quad F_{\text{freq}} = D_{\text{freq}}(F_{\text{in}}) \tag{1}$$

where both F_{spatial} and F_{freq} are lists containing feature maps at multiple scales. We denote the feature map at a specific layer k as F^k_{spatial} and F^k_{freq}.

Specifically, while both decoders follow a similar hierarchical upsampling structure, their core computational blocks are distinct. The spatial branch uses the WideResNet50 [16] decoder, which is pre-trained on ImageNet and kept frozen during training. Conversely, the frequency decoder D_{freq} is constructed from customized residual blocks. Inside each block, the input feature channels are split into a spatial part X_s and a frequency part X_f. These parts are processed in parallel to generate intermediate representations H_s and H_f:

$$H_s = \text{Conv}_\text{s}(X_s), \quad H_f = \mathcal{F}^{-1}\left(\text{Conv}_\text{f}\left(\mathcal{F}(X_f)\right)\right), \tag{2}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ denote the FFT and its inverse, while Conv_s and Conv_f represent the convolutional operations in the spatial and frequency paths, respectively. To adaptively integrate these dual-domain features, we compute a local mixing map M via a 1×1 convolution followed by a Sigmoid activation:

$$M = \sigma\left(\text{Conv}_{1\times 1}(H_s \oplus H_f)\right), \tag{3}$$

where $\oplus$ denotes channel concatenation. The final output of the block, Y_{block}, is then computed as a dynamically weighted sum:

$$Y_{\text{block}} = M \odot H_s + (1 - M) \odot H_f. \tag{4}$$

This design allows the network to intrinsically balance fine-grained spatial details and global frequency patterns pixel-by-pixel without relying on complex external fusion modules.

3) Dynamic Gated Fusion (DGF): We introduce a DGF mechanism to synthesize the final representations from the two student branches. Unlike the local mixing in OFS, DGF operates at a macro-level on the highest-level features (layer k). We first generate a spatially-aware gating map W_{raw} via a convolutional block followed by a Sigmoid activation:

$$W_{\text{raw}} = \sigma\left(\text{Conv}(F^k_{\text{spatial}} \oplus F^k_{\text{freq}})\right), \tag{5}$$

where $\oplus$ indicates channel concatenation. To introduce a global inductive bias, we incorporate a scalar hyperparameter $\lambda_{\text{bias}} \in [0, 1]$, which acts as a controller to shift the fusion focus. The final adjusted gating map W_g is computed as:

$$W_\text{g} = W_{\text{raw}} \odot \lambda_{\text{bias}} + (1 - W_{\text{raw}}) \odot (1 - \lambda_{\text{bias}}), \tag{6}$$

where $\odot$ denotes element-wise multiplication. The fusion is performed hierarchically. The dynamic gate W_g is applied exclusively to the highest-level features to perform a biased weighted summation:

$$F^k_{\text{fused}} = W_\text{g} \odot F^k_{\text{spatial}} + (1 - W_\text{g}) \odot F^k_{\text{freq}}. \tag{7}$$

For all lower-level features (layer $j < k$), we employ a simple fixed averaging strategy to ensure the stable integration of fine-grained details.

$$F^j_{\text{fused}} = \frac{1}{2} \cdot (F^j_{\text{spatial}} + F^j_{\text{freq}}). \tag{8}$$

The final output is the collection $F_{\text{fused}} = \{F^1_{\text{fused}}, \dots, F^k_{\text{fused}}\}$.

B. Uncertainty-Guided Distillation (UGD)

We employ an Uncertainty-Guided Distillation to prevent the student from learning the inherent uncertainty of the teacher network. We use Monte Carlo Dropout to quantify the uncertainty. By performing multiple forward propagations with dropout enabled during training, we use the mean of the predictions $\hat{\mu}_k(I)$ as the final output of the teacher network. $\hat{\mu}_k(I)$ is the average feature map from the k-th layer for input I, and the number of MC forward propagations is T. The set of outputs from the k-th layer is $\{\phi_k^1, \dots, \phi_k^T\}$, where each element $\phi_k^t(I)$ is the feature map from the t-th stochastic forward pass. The uncertainty map $U_k(I)$ is then defined as the sample variance at each position (h, w):

$$U_k(I)_{h,w} = \frac{1}{T-1}\sum_{t=1}^{T}\left(\phi_k^t(I)_{h,w} - \hat{\mu}_k(I)_{h,w}\right)^2. \quad (9)$$

Based on equation 9, we define an inverse uncertainty weight map $W_{\text{u}}^k(I)$, where the subscript u stands for uncertainty:

$$W_{\text{u}}^k(I) = \frac{1}{1+U_k(I)}. \quad (10)$$

This weight map assigns lower values to areas of high uncertainty. Our weighting strategy is grounded in Bayesian modeling of aleatoric uncertainty. Mathematically, this corresponds to maximizing the Gaussian log-likelihood, where the reconstruction loss is inherently attenuated by the inverse variance. By weighting the distance with $W_{\text{u}}^k(I) \propto 1/(1+U_k(I))$, we effectively down-weight regions where the teacher is uncertain (e.g., ambiguous boundaries or complex textures), preventing the student from learning noise. Furthermore, we utilize Cosine distance rather than L_2 distance because it focuses on vector direction (semantic content) rather than magnitude, making the objective more robust to global illumination changes common in industrial settings. The total loss for each layer $\mathcal{L}_{\text{mc}}^k$ is now defined as the spatial average of the element-wise weighted cosine distance, plus a regularization term:

$$\begin{aligned}\mathcal{L}_{\text{mc}}^k = \mathbb{E}_{h,w}\left[\left(1 - \frac{\hat{\mu}_k(I)\cdot\psi^k(I)}{||\hat{\mu}_k(I)||_2 \cdot ||\psi^k(I)||_2}\right) \odot W_u^k(I)\right] \\ + \lambda_\mu \cdot \mathbb{E}_{h,w}[U_k(I)],\end{aligned} \quad (11)$$

where $\mathbb{E}_{h,w}[\cdot]$ denotes the mean over spatial dimensions, $\psi^k(I)$ is the feature map of the student model, and λ_μ is a hyperparameter for the uncertainty regularization. Assume that the feature mean output by the teacher network is $\hat{\mu} = \{\hat{\mu}_1, \dots, \hat{\mu}_k\}$, the uncertainty map is $U = \{U_1, \dots, U_k\}$, the final uncertainty-guided loss $\mathcal{L}_{\text{uncertainty}}(\psi, \mu, U)$ is the sum over all N selected layers:

$$\mathcal{L}_{\text{uncertainty}} = \sum_{k=1}^{N}\mathcal{L}_{\text{mc}}^k. \quad (12)$$

This comprehensive loss $\mathcal{L}_{\text{uncertainty}}$ forms the basis of our end-to-end optimization strategy. We apply this loss function directly to the final fused features F_{fused}. Therefore, the total network loss $\mathcal{L}_{\text{total}}$ is defined as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{uncertainty}}(F_{\text{fused}}, \hat{\mu}, U), \quad (13)$$

where the student features ψ^k used internally in the calculation of $\mathcal{L}_{\text{uncertainty}}$ are now explicitly the fused features F^k_{fused}. This single objective jointly optimizes all network components, and while the individual branch losses are monitored for analysis, they do not contribute to the gradient updates.

IV. Experiments

A. Experimental Setups

1) Datasets: We evaluate our method on two benchmarks: HSS-IAD and MVTec AD. The HSS-IAD dataset aggregates 8,580 images from industrial subsets including Casting [23], KolektorSDD [24], KolektorSDD2 [25], MTD [26], and STEEL [27], covering both object and texture anomalies. The MVTec AD benchmark comprises 5,354 high-resolution images across 15 categories. For both, training utilizes exclusively anomaly-free samples.

2) Hyperparameter Setting: We train a separate model for each category with images resized to 256×256 and normalized via ImageNet statistics. The network is optimized using Adam ($lr = 1\times10^{-3}$) with a batch size of 16 on a single NVIDIA RTX A6000. For uncertainty estimation, we perform 6 Monte Carlo Dropout forward passes, which provides a reliable estimate of teacher uncertainty; fewer passes (e.g., 2–4) yield unstable uncertainty maps, while more passes (e.g., >8) offer diminishing accuracy gains at the cost of increased training overhead. The gated fusion bias λ_{bias} is set to 0.4 to prioritize the frequency branch for texture sensitivity.

3) Evaluation Metrics: Performance is evaluated using image-level AUC (I-AUC) for classification, and three pixel-level metrics for localization: P-AUC, area under the per-region overlap curve (PRO), and Average Precision (P-AP).

B. Experiment Results

Tables I and II detail the comparative results on the HSS-IAD and MVTec AD datasets, respectively, where the optimal and sub-optimal results are bolded and underlined. On the HSS-IAD dataset, our method improves upon the state-of-the-art by **4.2%** in I-AUC, **1.5%** in P-AUC, and **0.67%** in PRO, demonstrating excellent performance in small-area anomaly classification and localization. On the MVTec AD dataset, our method achieves highly competitive performance and surpasses leading methods on average metrics. Anomaly heatmaps are visualized in Figures 3 and 4 (top to bottom: test image, ground truth, anomaly heatmap).

C. Ablation Studies

To verify the contribution of our proposed components, we performed a detailed ablation study, with the results presented in Table III. The study reveals that our component designs contribute positively to all performance metrics. Among them, the OFS module is particularly effective, providing the most substantial individual performance boost. This confirms that

TABLE I
ANOMALY DETECTION METRICS (I-AUC/P-AUC/PRO/P-AP) ON THE HSS-IAD DATASET(%).

Category	DesTSeg [17]	SimpleNet [18]	DRÆM [19]	UniAD [20]	RD [2]	RD++ [3]	Dinomaly [21]	**Ours**
MTD	95.7/-/67.8/30.2	57.4/-/24.9/6.0	79.0/-/46.5/16.2	**98.8**/-/**74.7**/27.9	96.4/-/70.7/27.6	92.3/70.0/74.3/**54.5**	98.5/-/68.3/44.6	98.79/**76.83**/74.58/53.54
STEEL	65.7/-/24.8/20.5	45.8/-/17.2/3.4	55.6/-/21.6/12.8	54.8/-/38.7/24.5	57.1/-/46.4/14.0	59.6/81.7/**56.1**/**33.6**	**66.4**/-/43.5/31.6	65.83/**83.59**/55.88/31.90
KolektorSDD	77.7/-/49.4/19.6	56.7/-/23.7/0.8	61.1/-/43.6/2.0	83.3/-/47.8/7.8	76.4/-/66.8/12.1	88.1/**90.7**/75.1/**27.4**	**95.3**/-/**94.4**/23.0	88.54/89.34/73.51/26.31
KolektorSDD2	95.2/-/89.9/71.4	54.2/-/15.4/1.0	79.5/-/58.2/38.9	96.1/-/94.2/46.0	**96.2**/-/94.5/49.9	95.5/98.5/95.2/**73.9**	95.2/-/91.6/70.1	95.60/**98.65**/**95.66**/73.43
Casting_C1	61.0/-/16.7/0.4	49.1/-/22.6/0.4	79.4/-/41.1/0.9	51.8/-/11.3/0.4	46.1/-/43.4/2.1	64.0/78.1/48.9/**13.7**	75.0/-/43.7/3.7	**81.58**/**84.13**/**54.00**/10.45
Casting_C2	50.9/-/47.6/1.1	49.1/-/19.1/0.3	58.7/-/46.9/1.0	50.6/-/34.3/0.7	64.9/-/62.3/1.9	73.1/87.7/68.7/17.4	58.8/-/46.9/2.6	**82.10**/**88.28**/**69.30**/**17.56**
Casting_C3	72.2/-/31.3/0.4	58.5/-/12.4/0.1	72.9/-/39.5/0.5	50.5/-/22.2/0.3	62.6/-/50.6/0.7	**73.7**/**83.8**/**57.4**/6.89	67.2/-/26.8/0.6	73.43/80.48/55.67/**7.76**
Average	74.1/-/46.8/20.5	53.0/-/19.3/1.7	69.5/-/42.5/10.3	69.4/-/46.2/15.4	71.4/-/62.1/15.5	78.0/84.4/67.7/**32.5**	79.5/-/59.3/25.2	**83.70**/**85.90**/**68.37**/31.56

TABLE II
ANOMALY DETECTION METRICS (I-AUC/P-AUC/PRO) ON THE MVTEC AD DATASET(%).

Category	PatchCore [14]	CFA [22]	RD [2]	RD++ [3]	**Ours**
Carpet	98.7 / 99.0 / 96.6	97.3 / **99.28** / 96.54	98.9 / 98.9 / 97.0	**100** / 99.2 / 97.7	**100** / 99.24 / **97.85**
Grid	98.2 / 98.7 / 96.0	99.2 / 98.12 / 94.04	**100** / 99.3 / 97.6	**100** / 99.3 / 97.7	**100** / **99.34** / **97.89**
Leather	**100** / 99.3 / 98.9	**100** / 99.37 / 97.43	**100** / 99.4 / 99.1	**100** / 99.4 / 99.2	**100** / **99.44** / **99.25**
Tile	98.7 / 95.4 / 87.3	99.4 / 95.21 / 89.26	99.3 / 95.6 / 90.6	**99.7** / **96.6** / 92.4	99.57 / 96.53 / **93.49**
Wood	99.2 / 95.0 / 89.4	**99.7** / 91.53 / 90.54	99.2 / 95.3 / 90.9	99.3 / **95.8** / 93.3	99.39 / 95.58 / **93.73**
Bottle	**100** / 98.6 / 96.2	**100** / **98.84** / 95.76	**100** / 98.7 / 96.6	**100** / 98.8 / 97.0	**100** / 98.81 / **97.18**
Cable	99.5 / 98.4 / 92.5	**99.8** / **99.87** / 94.17	95.0 / 97.4 / 91.0	99.2 / 98.4 / 93.9	99.38 / 98.48 / **94.64**
Capsule	98.1 / 98.8 / 95.5	97.3 / **99.11** / 93.66	96.3 / 98.7 / 95.8	**99.0** / 98.8 / **96.4**	98.44 / 98.62 / 96.31
Hazelnut	**100** / 98.7 / 93.8	**100** / 98.85 / 95.75	99.9 / 98.9 / 95.5	**100** / **99.2** / 96.3	**100** / 99.18 / **96.57**
Metal nut	**100** / 98.4 / 91.4	**100** / **99.15** / **94.54**	**100** / 97.3 / 92.3	**100** / 98.1 / 93.0	**100** / 98.17 / 93.45
Pill	96.6 / 97.4 / 93.2	97.9 / **98.93** / **97.19**	96.6 / 98.2 / 96.4	98.4 / 98.3 / 97.0	**98.56** / 98.03 / 96.61
Screw	98.1 / 99.4 / 97.9	97.3 / 98.91 / 95.23	97.0 / 99.6 / 98.2	**98.9** / **99.7** / **98.6**	98.81 / 99.57 / 98.30
Toothbrush	**100** / 98.7 / 91.5	**100** / 98.96 / 91.14	99.5 / 99.1 / 94.5	**100** / 99.1 / 94.2	**100** / **99.17** / **95.04**
Transistor	**100** / 96.3 / 83.7	**100** / **98.06** / **95.35**	96.7 / 92.5 / 78.0	98.5 / 94.3 / 81.8	99.63 / 95.33 / 87.91
Zipper	99.4 / 98.8 / **97.1**	**99.6** / **99.02** / 95.95	98.5 / 98.2 / 95.4	98.6 / 98.8 / 96.3	99.19 / 98.97 / 96.82
Average	99.1 / 98.06 / 93.4	99.17 / 98.15 / 94.44	98.46 / 97.81 / 93.93	99.44 / 98.25 / 94.99	**99.53** / **98.30** / **95.67**

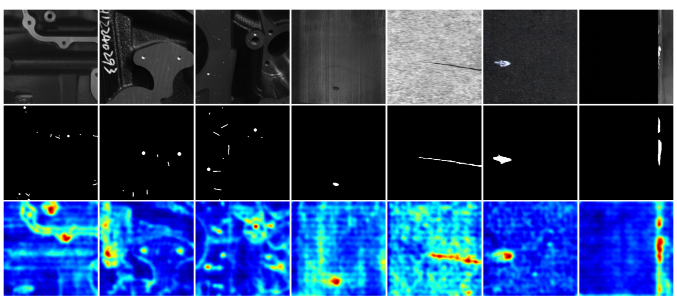

Fig. 3. Visualization of our method on the HSS-IAD dataset (From left to right: Casting_C1, C2, C3, MTD, SDD, SDD2, STEEL).

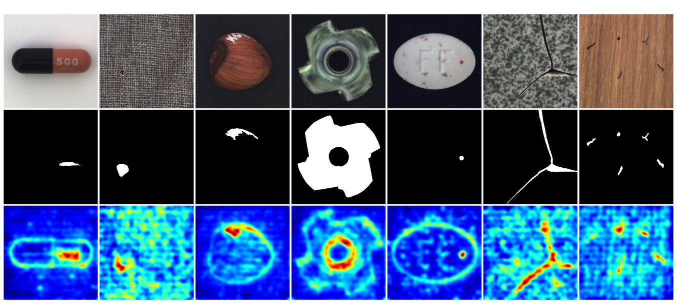

Fig. 4. Visualization of our method on the MVTec AD dataset (From left to right: Capsule, Carpet, Hazelnut, Metal Nut, Pill, Tile, Wood).

TABLE III
ABLATION STUDY OF OUR MODULES ON HSS-IAD DATASET(%).

CEB	**OFS**	**UGD**	**I-AUC**	**P-AUC**	**PRO**
			78.03	84.36	67.66
✓			78.70	84.71	67.73
	✓		79.52	84.96	68.25
		✓	78.88	84.51	68.20
✓	✓		81.38	84.31	68.36
✓	✓	✓	**83.70**	**85.90**	**68.37**

our architectural designs are not redundant and that each component is essential for achieving superior performance.

V. CONCLUSION

In this paper, we have proposed a Dual-View Distillation Framework for industrial anomaly detection. By synergizing an Omni-Frequency Student with Monte Carlo Dropout for uncertainty estimation, our method enforces global structural consistency while guiding the student to learn only from confident teacher predictions, effectively mitigating over-generalization. Furthermore, we integrate a Cross-scale Efficiency Block to synthesize multi-scale features, preserving salient fine-grained details often lost in standard bottlenecks. Extensive experiments on challenging datasets validate that our approach excels in identifying subtle defects and handling ambiguous industrial scenarios.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China under Grant 61871186.

REFERENCES

[1] G. Wang, S. Han, E. Ding, and D. Huang, "Student-teacher feature pyramid matching for anomaly detection," in *32nd British Machine Vision Conference 2021, BMVC 2021, Online, November 22-25, 2021*. BMVA Press, 2021, p. 306. [Online]. Available: https://www.bmvc2021-virtualconference.com/assets/papers/1273.pdf

[2] H. Deng and X. Li, "Anomaly detection via reverse distillation from one-class embedding," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*. IEEE, 2022, pp. 9727–9736. [Online]. Available: https://doi.org/10.1109/CVPR52688.2022.00951

[3] T. D. Tien, A. T. Nguyen, N. H. Tran, T. D. Huy, S. T. M. Duong, C. D. T. Nguyen, and S. Q. H. Truong, "Revisiting reverse distillation for anomaly detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*. IEEE, 2023, pp. 24 511–24 520. [Online]. Available: https://doi.org/10.1109/CVPR52729.2023.02348

[4] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, and R. Garnett, Eds., 2017, pp. 5574–5584. [Online]. Available: https://proceedings.neurips.cc/paper/2017/hash/2650d6089a6d640c5e85b2b88265dc2b-Abstract.html

[5] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Mvtec AD - A comprehensive real-world dataset for unsupervised anomaly detection," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*. Computer Vision Foundation / IEEE, 2019, pp. 9592–9600. [Online]. Available: http://openaccess.thecvf.com/content_CVPR_2019/html/Bergmann_MVTec_AD_--_A_Comprehensive_Real-World_Dataset_for_Unsupervised_Anomaly_CVPR_2019_paper.html

[6] Q. Wang, S. Gao, J. Hu, J. Yu, X. Tong, Y. Li, and W. Zhang, "HSS-IAD: A heterogeneous same-sort industrial anomaly detection dataset," *CoRR*, vol. abs/2504.12689, 2025. [Online]. Available: https://doi.org/10.48550/arXiv.2504.12689

[7] J. Liu, G. Xie, J. Wang, S. Li, C. Wang, F. Zheng, and Y. Jin, "Deep industrial image anomaly detection: A survey," *Mach. Intell. Res.*, vol. 21, no. 1, pp. 104–135, 2024. [Online]. Available: https://doi.org/10.1007/s11633-023-1459-z

[8] Y. Cui, Z. Liu, and S. Lian, "A survey on unsupervised anomaly detection algorithms for industrial images," *IEEE Access*, vol. 11, pp. 55 297–55 315, 2023.

[9] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," in *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2019, Volume 5: VISAPP, Prague, Czech Republic, February 25-27, 2019*, A. Trémeau, G. M. Farinella, and J. Braz, Eds. SciTePress, 2019, pp. 372–380. [Online]. Available: https://doi.org/10.5220/0007364503720380

[10] Y. Liang, J. Zhang, S. Zhao, R. Wu, Y. Liu, and S. Pan, "Omni-frequency channel-selection representations for unsupervised anomaly detection," *IEEE Trans. Image Process.*, vol. 32, pp. 4327–4340, 2023. [Online]. Available: https://doi.org/10.1109/TIP.2023.3293772

[11] S. Zhao, A. Liu, D. Qu, and S. Zhao, "Few-shot anomaly detection based on diffusion generation model and multi-scale feature manipulation," in *International Joint Conference on Neural Networks, IJCNN 2025, Rome, Italy, June 30 - July 5, 2025*. IEEE, 2025, pp. 1–8. [Online]. Available: https://doi.org/10.1109/IJCNN64981.2025.11227359

[12] C. Hu, K. Chen, and H. Shao, "A semantic-enhanced method based on deep SVDD for pixel-wise anomaly detection," in *2021 IEEE International Conference on Multimedia and Expo, ICME 2021, Shenzhen, China, July 5-9, 2021*. IEEE, 2021, pp. 1–6. [Online]. Available: https://doi.org/10.1109/ICME51207.2021.9428370

[13] J. Lei, X. Hu, Y. Wang, and D. Liu, "Pyramidflow: High-resolution defect contrastive localization using pyramid normalizing flow," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*. IEEE, 2023, pp. 14 143–14 152. [Online]. Available: https://doi.org/10.1109/CVPR52729.2023.01359

[14] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. V. Gehler, "Towards total recall in industrial anomaly detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*. IEEE, 2022, pp. 14 298–14 308. [Online]. Available: https://doi.org/10.1109/CVPR52688.2022.01392

[15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[16] S. Zagoruyko and N. Komodakis, "Wide residual networks," in *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016*, R. C. Wilson, E. R. Hancock, and W. A. P. Smith, Eds. BMVA Press, 2016. [Online]. Available: https://bmva-archive.org.uk/bmvc/2016/papers/paper087/index.html

[17] X. Zhang, S. Li, X. Li, P. Huang, J. Shan, and T. Chen, "Destseg: Segmentation guided denoising student-teacher for anomaly detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*. IEEE, 2023, pp. 3914–3923. [Online]. Available: https://doi.org/10.1109/CVPR52729.2023.00381

[18] Z. Liu, Y. Zhou, Y. Xu, and Z. Wang, "Simplenet: A simple network for image anomaly detection and localization," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*. IEEE, 2023, pp. 20 402–20 411. [Online]. Available: https://doi.org/10.1109/CVPR52729.2023.01954

[19] V. Zavrtanik, M. Kristan, and D. Skocaj, "Dræm - A discriminatively trained reconstruction embedding for surface anomaly detection," in *2021 IEEE/CVF International Conference on Computer Vision, ICCV 2021, Montreal, QC, Canada, October 10-17, 2021*. IEEE, 2021, pp. 8310–8319. [Online]. Available: https://doi.org/10.1109/ICCV48922.2021.00822

[20] Z. You, L. Cui, Y. Shen, K. Yang, X. Lu, Y. Zheng, and X. Le, "A unified model for multi-class anomaly detection," in *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, Eds., 2022. [Online]. Available: http://papers.nips.cc/paper_files/paper/2022/hash/1d774c112926348c3e25ea47d87c835b-Abstract-Conference.html

[21] J. Guo, S. Lu, W. Zhang, F. Chen, H. Li, and H. Liao, "Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*. Computer Vision Foundation / IEEE, 2025, pp. 20 405–20 415. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2025/html/Guo_Dinomaly_The_Less_Is_More_Philosophy_in_Multi-Class_Unsupervised_Anomaly_CVPR_2025_paper.html

[22] S. Lee, S. Lee, and B. C. Song, "Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization," *IEEE Access*, vol. 10, pp. 78 446–78 454, 2022.

[23] K. Mao, P. Wei, Y. Wang, M. Liu, S. Wang, and N. Zheng, "CSDD: A benchmark dataset for casting surface defect detection and segmentation," *IEEE CAA J. Autom. Sinica*, vol. 12, no. 5, pp. 947–960, 2025. [Online]. Available: https://doi.org/10.1109/JAS.2025.125228

[24] D. Tabernik, S. Sela, J. Skvarc, and D. Skocaj, "Segmentation-based deep-learning approach for surface-defect detection," *J. Intell. Manuf.*, vol. 31, no. 3, pp. 759–776, 2020. [Online]. Available: https://doi.org/10.1007/s10845-019-01476-x

[25] J. Bozic, D. Tabernik, and D. Skocaj, "Mixed supervision for surface-defect detection: From weakly to fully supervised learning," *Comput. Ind.*, vol. 129, p. 103459, 2021. [Online]. Available: https://doi.org/10.1016/j.compind.2021.103459

[26] Y. Huang, C. Qiu, and K. Yuan, "Surface defect saliency of magnetic tile," *Vis. Comput.*, vol. 36, no. 1, pp. 85–96, 2020. [Online]. Available: https://doi.org/10.1007/s00371-018-1588-5

[27] A. Grishin, BorisV, iBardintsev, inversion, and Oleg, "Severstal: Steel defect detection," https://kaggle.com/competitions/severstal-steel-defect-detection, 2019, kaggle.