CDAD: A Multi-Scale Feature Fusion Model for Cross-Domain Anomaly Detection

Hui Zhang[1], Hongyu Wang[1], Jianyong Zhu[1,3], Hongtao Wang[1,2,†]

[1]*Department of Computer, North China Electric Power University*

[2]*Engineering Research Center of Intelligent Computing for Complex Energy Systems, Ministry of Education*

[3]*Hebei Key Laboratory of Knowledge Computing for Energy & Power*

{*zh, why, zhujy, wanght*}*@ncepu.edu.cn*

***Abstract*— Anomaly detection aims to identify deviations by learning normal data distributions, mainly covering industrial, semantic, and medical domains. Most existing methods, however, are domain-specific and degrade when transferred. To address this limitation, we proposed CDAD, a unified Cross-Domain Anomaly Detection model. CDAD leverages a pre-trained feature extractor with a feature adaptor to align cross-domain features and tighten the decision boundary of normal samples. A hybrid loss emphasizes hard-to-classify anomalies, while a Multi-Scale Feature Fusion Discriminator integrates global and local features. Extensive experiments on five datasets across three domains demonstrate that CDAD consistently outperforms recent state-of-the-art approaches, highlighting its robustness and effectiveness in cross-domain anomaly detection. The code is available at https://github.com/Stardust457/CDAD.**

***Index Terms*—Cross-Domain, Anomaly Detection, Multi-Scale Feature Fusion, Anomaly Localization, Focal Loss**

I. Introduction

The objective of anomaly detection lies in learning the distribution of normal samples within a given dataset to precisely identify data deviating from this distribution. This technology finds extensive applications across various domains, spanning industrial anomaly detection, semantic anomaly detection, and medical anomaly detection. Specifically, it can be deployed in industrial production lines to identify defects such as scratches, dents, and contamination on mechanical components. In autonomous driving scenarios, the detection of previously unseen objects constitutes a paradigm of semantic anomaly detection [1]. When applied to the medical domain [2], [3], anomaly detection techniques enable identification of subtle pathological lesions that are difficult to perceive visually, providing critical references for clinical diagnosis.

A defining characteristic of anomaly detection is the scarcity of anomalous training samples. Consequently, models are trained exclusively on normal data, requiring them to generalize to entirely unseen anomaly types during inference. This challenging setting has motivated substantial methodological innovation. Recent approaches [4]–[8] leverage pre-trained feature extractors and advanced generative models to capture semantic anomalies effectively. In parallel, specialized architectures [9]–[12] have been developed for industrial scenarios, achieving impressive results in both defect detection and precise localization.

†Corresponding author.

Despite these advances, a critical and often overlooked limitation persists: the prevailing research paradigm is inherently domain-specific. Existing methods are typically designed and optimized for a single domain, leading to fragmented solutions that fail to generalize across different application contexts. This fragmentation poses significant practical challenges: it necessitates the development, maintenance, and deployment of distinct models for different tasks, increasing complexity and resource overhead. More fundamentally, it reveals a deep-seated trade-off in current anomaly detection paradigms. For instance, SimpleNet [10], the former state-of-the-art on the industrial MVTec-AD benchmark [13], experiences severe performance degradation on the semantic Aircraft-FGVC dataset [14]. The underlying cause is illustrative: SimpleNet's patch-level defect detection excels at capturing localized structural anomalies but lacks the global semantic modeling necessary for semantic tasks. Conversely, methods designed for semantic anomaly detection [4], [8] effectively learn holistic semantic representations yet struggle with fine-grained local defects, rendering them unsuitable for industrial applications. Even GeneralAD [15], explicitly designed for cross-domain detection, shows degraded performance on medical datasets due to its inability to adequately emphasize hard-to-classify anomalies.

These observations underscore a fundamental dilemma: methods optimized for local structural defects sacrifice global semantic understanding, while those emphasizing semantic coherence overlook fine-grained irregularities. This dilemma not only limits the practical applicability of any single model but also hinders the development of robust, general-purpose anomaly detection systems. The significant disparity in data characteristics and anomaly types across domains—from textured surfaces in manufacturing to natural objects in autonomous driving and subtle lesions in medical imaging—makes cross-domain generalization exceptionally challenging. Yet, the ability to handle such diversity with a single, adaptable model is of paramount practical importance. It aligns with the vision of developing foundational models for visual perception that can be efficiently applied to downstream tasks with minimal domain-specific tuning. Therefore, moving beyond domain-specific optimization towards unified cross-domain anomaly detection is not merely an incremental improvement but a necessary step to enhance the scalability, robustness, and practical utility of anomaly detection systems.

To overcome this limitation and address the urgent need for generalization, we propose CDAD (Cross-Domain Anomaly Detection), a unified framework that harmonizes both local and global anomaly detection capabilities. Our approach introduces three key innovations. First, we design a Multi-Scale Feature Fusion Discriminator (MFFD) that simultaneously models comprehensive global semantic relationships between image patches and identifies localized structural defects, enabling effective detection across semantic, logical, and structural anomaly types. Second, we develop a feature adaptation strategy that tightens the decision boundaries of normal samples by projecting pre-trained features into a domain-aligned target space, significantly enhancing out-of-distribution detection. Third, recognizing that standard loss functions treat all samples equally—causing models to overlook challenging anomalies—we propose a hybrid loss combining binary cross-entropy and focal loss [16]. This design strategically amplifies the contribution of hard-to-classify anomalies, compelling the model to focus on the most informative cases.

We validate the cross-domain efficacy of CDAD through comprehensive experiments across five benchmark datasets spanning three highly dissimilar domains: MPDD [17], MVTec-AD [13], and MVTec-LOCO [18] for industrial anomaly detection; Aircraft-FGVC [14] for semantic anomaly detection; and APTOS [19] for medical anomaly detection. Our results demonstrate that CDAD achieves state-of-the-art performance on MPDD and MVTec-LOCO while maintaining competitive excellence on MVTec-AD, Aircraft-FGVC, and APTOS, confirming its superior cross-domain generalization capability.

Our main contributions are summarized as follows:

- We propose CDAD, a unified cross-domain anomaly detection framework featuring a Multi-Scale Feature Fusion Discriminator (MFFD) that effectively integrates global semantic modeling with local feature extraction, enabling comprehensive and domain-agnostic detection of semantic, logical, and structural anomalies.
- We introduce a feature adaptor module that tightens normal sample decision boundaries by mapping pre-trained features to a target feature space, alongside a hybrid loss function combining binary cross-entropy and focal loss to emphasize hard-to-classify anomalies.
- Through extensive evaluation on five datasets across three domains, we demonstrate that CDAD achieves state-of-the-art or highly competitive performance compared to recent methods, exhibiting robust cross-domain generalization.

II. Related Work

Anomaly detection encompasses diverse anomaly patterns across multiple domains, where each category typically necessitates tailored solutions. In this section, we present a comprehensive analysis of three critical scenarios relevant to our work, along with their corresponding methodological approaches.

A. Industrial and Medical Anomaly Detection

Industrial anomaly detection requires fine-grained identification of subtle defects. Existing approaches are broadly categorized into synthesis-based [20]–[22] and embedding-based methods [2], [9], [11], [23]–[26]. The former synthesizes anomalies (e.g., via diffusion [22]), while the latter often uses teacher-student frameworks to learn precise distribution boundaries [9]. These techniques have been extended to medical data [2], which shares the need for localized, fine-grained detection but introduces greater challenges due to diverse and subtle lesions.

A common limitation of current models is their focus on patch-level anomalies while overlooking inter-patch contextual relationships. This restricts their ability to capture global semantic features, which are essential for semantic anomaly detection. In contrast, our CDAD model jointly localizes fine-grained defects and models global semantic correlations, offering a unified cross-domain detection framework.

B. Semantic Anomaly Detection

Semantic anomaly detection aims to identify samples that deviate from the training data distribution at the semantic level. When a significant distributional deviation exists between normal and anomalous samples in a dataset, as exemplified by CIFAR-10 [27], this scenario falls under the category of semantic anomaly detection. The more challenging near semantic anomaly detection occurs when normal and anomalous samples belong to the same category with minimal semantic discrepancies, such as in Aircraft-FGVC [14]. Current approaches primarily include leveraging features from pre-trained extractor [4], [8], [28] and self-supervised learning [7], [29], [30] methods. While these methods demonstrate strong performance on semantic datasets, their insufficient capability in local defect detection leads to suboptimal results on industrial benchmarks.

C. Cross-Domain Anomaly Detection

Cross-domain anomaly detection aims to establish a universal framework for identifying structural, logical, and semantic anomalies across heterogeneous domains. Method [31] bridges industrial and medical domains through vision-language models (VLMs), while method [32] introduces a generalized few-shot detection paradigm via training-free universal models, unifying anomaly detection across industrial, logical, and medical domains. Methods [33], [34] are dedicated to open-set anomaly detection, demonstrating exceptional performance across nine datasets from three domains. GeneralAD [15] further integrates industrial anomaly detection, semantic anomaly detection, and near anomaly detection. However, GeneralAD exhibits performance degradation on medical datasets. This limitation arises from its single loss function design, causing the model to treat normal and anomalous samples equally. This makes it difficult to handle the numerous hard-to-classify anomalies in medical datasets.

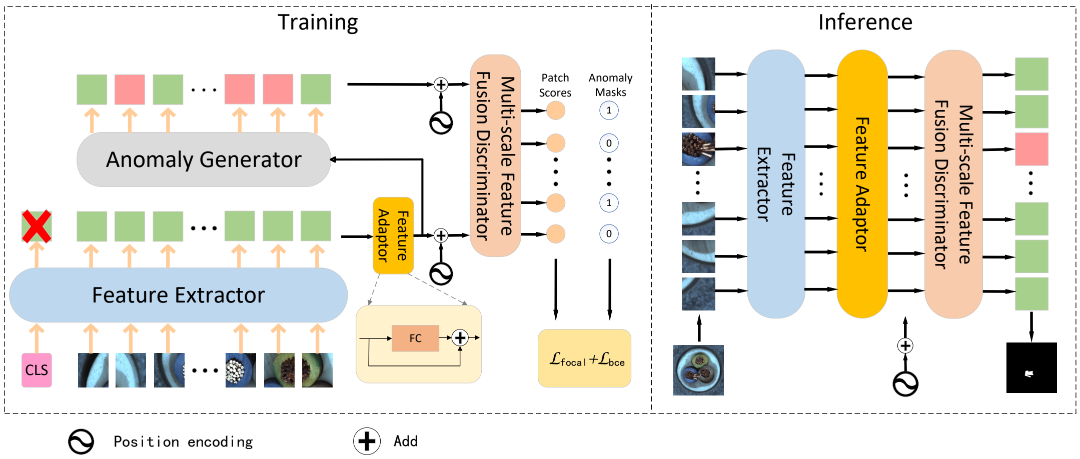

Fig. 1. **Model overview of CDAD.** The CDAD comprises four key components: a feature extractor, a feature adaptor, an anomaly generator, and the MFFD module, accompanied by a hybrid loss function.

III. Method

The proposed CDAD consists of four components: a feature extractor, a feature adaptor, an anomaly generator, and the MFFD module, trained with a hybrid loss (Figure 1). Using state-of-the-art pretrained models [35], [36], the extractor $\mathcal{E}$ maps an input $x \in \mathbb{R}^{3\times H\times W}$ to $y = \mathcal{E}(x)$. The adaptor $\mathcal{Q}$ produces $z = \mathcal{Q}(y)$, which is processed by MFFD to obtain the anomaly score $\mathcal{D}(z)$. The decision function $\mathcal{F}$ is defined as:

$$\mathcal{F}(x, \mathcal{E}, \mathcal{Q}, \mathcal{D}) = \begin{cases} 0 & \text{if } x \sim \mathcal{P}_{\text{Normal}}, \\ 1 & \text{otherwise}, \end{cases} \tag{1}$$

where 0 denotes normal samples and 1 anomalous ones.

During training, normal features are first processed by the feature adaptor to obtain adapted representations. These adapted features are then fed into the anomaly generator to synthesize anomaly features. Both adapted and anomaly features, with positional encodings, are subsequently input into MFFD to produce patch-level anomaly scores and compute loss with anomaly masks. During inference, the anomaly generator is removed. Details of each CDAD component are given below.

A. Feature Extractor

CDAD uses a Vision Transformer pretrained with DINOv2 [36] as feature extractor. An image $x \in \mathbb{R}^{3\times H\times W}$ is divided into $N = \frac{H}{P} \times \frac{W}{P}$ patches of size $P\times P$, concatenated with a CLS token, and encoded by $\mathcal{E}$. We extract N attention values from the i-th attention head, excluding the CLS token, denoted as $\mathcal{M}_i(x) = [m_1^i, \ldots, m_N^i]$. The encoder output $\mathcal{E}(x) = [e_1, \ldots, e_N]$ excludes CLS, as our task requires cross-domain generalization. $\mathcal{M}_i$ is used by the anomaly generator to create logical anomalies.

B. Feature Adaptor

CDAD unifies anomaly detection across domains. While most methods rely on extractors pretrained on natural datasets like ImageNet-1k [37], industrial and medical data often differ in distribution. To address this, a bias-free fully connected feature adaptor with residual connection maps extracted features $\mathcal{E}(x)$ to adapted features $\mathcal{Q}(y)$. Let $y = \mathcal{E}(x)$, then

$$\mathcal{Q}(y) = y \cdot W + y, \tag{2}$$

where W is the weight matrix.

C. Anomaly Generator

To enable cross-domain anomaly detection, we simulate diverse anomalies. Since image patch sequences exhibit semantic properties [35], [38], [39], generalization to unseen normals is possible, but scarce anomalies hinder boundary optimization. Prior works use diffusion models [6], which are costly and fail to capture structural or logical patterns. We instead propose an anomaly generator with Gaussian noise injection and attention shuffling. Given adapted features $\mathcal{Q}(y)$, abnormal features are

$$\widetilde{\mathcal{Q}}(y) = \mathcal{Q}(y) + \epsilon_\theta, \tag{3}$$

where ϵ_θ is Gaussian noise. Structural anomalies arise from perturbing random patches, semantic ones from all patches. Logical anomalies are simulated by shuffling the top K high-attention patches using attention head $\mathcal{M}_i$:

$$\widetilde{\mathcal{Q}}(y) = \text{shuffle}(\mathcal{Q}(y), \mathcal{M}_i(x)). \tag{4}$$

D. Multi-Scale Feature Fusion Discriminator

Inputs to the MFFD module are either normal or abnormal features. To ensure robustness across domains, MFFD captures semantic, structural, and logical anomalies. We use a single transformer encoder to model global semantics via self-attention [40]. Inspired by [41], a Conv module with 1×1 and 7×7 kernels extracts multi-scale local features. The

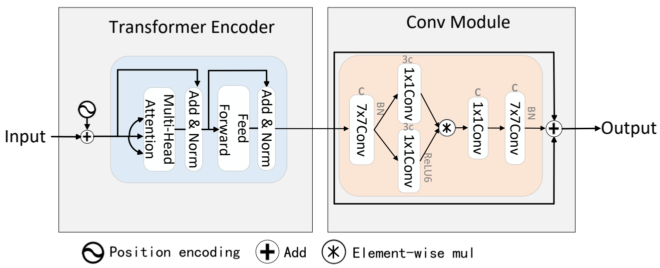

Fig. 2. **The proposed MFFD overview.** In the architecture diagram of MFFD, BN denotes Batch Normalization with c indicating the number of channels.

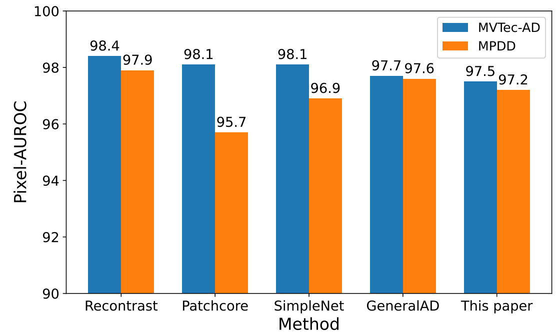

Fig. 3. **Anomaly localization results.** CDAD vs other baselines on MVTec-AD and MPDD (Pixel-AUROC).

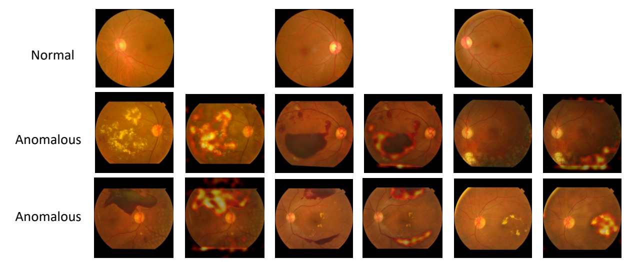

Fig. 4. **Medical visualization.** CDAD on APTOS.

Conv module is connected in series with the encoder, and their outputs are fused through two residual connections to obtain the final representation. This representation serves as the patch-level anomaly score, as illustrated in Figure 2.

E. Loss Function

In supervised anomaly detection, binary cross-entropy (BCE) loss is commonly used, but it treats easy and hard samples equally, causing rare difficult anomalies to be dominated by numerous easy normals. To emphasize challenging cases, we incorporate focal loss [16]:

$$\mathcal{L}_{focal} = -\alpha_t(1 - p_t)^\gamma \log(p_t), \tag{5}$$

where p_t is the predicted probability, γ controls focusing, and α_t balances classes. The final objective combines BCE and focal loss:

$$\mathcal{L} = \mathcal{L}_{bce} + \mathcal{L}_{focal}. \tag{6}$$

This design leverages BCE's stability while enhancing sensitivity to hard anomalies.

F. Inference

As anomaly sizes differ across domains, MFFD aggregates patch scores using a domain-specific K. For a test image x, patch-level scores $\text{Score}_P(i)$ are computed, and the image-level score is the mean of the top-K values:

$$\text{Score}(x) = \frac{1}{K}\sum_{i=1}^{K}\text{Score}_P(i), \tag{7}$$

where K is adaptively chosen per domain.

IV. Experiments

In this section, we present the datasets, baselines, implementation details, anomaly detection results, anomaly localization results, and ablation study.

A. Datasets and Baselines

Our experiments cover three domains: industrial (MVTec-AD [13], MVTec-LOCO [18], MPDD [17]), semantic (Aircraft-FGVC [14]), and medical (APTOS [19]) anomaly detection.

We compare CDAD with representative baselines, including PatchCore [11], KDAD [8], RDAD [26], Recontrast [2], SimpleNet [10], RealNet [22], and GeneralAD [15], covering industrial, semantic, and medical anomaly detection tasks.

B. Implementation Details

We adopt a pretrained DINOV2 [36] feature extractor, using features from its last layer. All inputs are resized to 518×518. Our feature adaptor is a bias-free fully-connected layer with a residual connection, trained at a learning rate of 10^{-4}. Added noise has a standard deviation of 0.25.

For industrial and medical datasets, random noise is applied; for MVTec-LOCO, attention shuffling generates logical anomalies. For near-anomaly detection, noise is added to all patches of an image. The MFFD is optimized with AdamW [42], a learning rate of 5×10^{-4}, weight decay 10^{-5}, and cosine annealing (decay factor 0.2). Focal loss parameters are $\gamma = 2$, $\alpha_t = 0.25$.

Training runs for 160 epochs on industrial/near-anomaly datasets and 20 epochs on medical datasets, all with batch size 16. The parameter K is set to 10 for industrial/medical data, and 1369 for near-anomaly detection, reflecting anomalous region sizes.

C. Anomaly Detection

We present CDAD results in Table I. CDAD shows strong generalization across domains: it surpasses prior SOTA on industrial datasets, outperforming GeneralAD [15] by 0.3% on MPDD and 0.6% on MVTec-LOCO, while matching SOTA on MVTec-AD. In semantic anomaly detection, CDAD ranks second to GeneralAD, and in medical datasets achieves 97.5% Image-AUROC, matching Recontrast [2].

CDAD benefits from MFFD for global–local feature capture, the feature adaptor for aligning pretrained features and

TABLE I

Anomaly detection results. We present a comparison of CDAD against other current SOTA methods based on the Image-AUROC metric, with all results rounded to one decimal place. The underlined data are reported by us.

	Industrial Anomaly Detection			Semantic Anomaly Detection	Medical Anomaly Detection
Method	MPDD	MVTec-LOCO	MVTec-AD	Aircraft-FGVC	APTOS
PatchCore	82.1	80.3	99.2	67.8	90.5
KDAD	72.1	67.5	85.8	85.8	-
RDAD	92.7	79.7	98.4	-	-
Recontrast	95.3	82.1	99.5	-	**97.5**
SimpleNet	94.8	77.6	**99.6**	83.6	93.4
RealNet	96.3	-	**99.6**	-	-
GeneralAD	97.8	84.9	99.2	**94.6**	96.1
CDAD	**98.1**	**85.5**	99.0	92.0	**97.5**

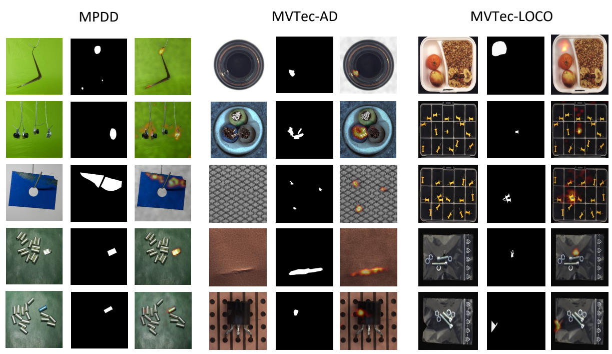

Fig. 5. **Industrial visualization.** CDAD on MPDD, MVTec-AD, and MVTec-LOCO.

compacting normal distributions, and the hybrid loss for emphasizing anomalies. Leveraging only synthesized anomalies in training, CDAD achieves superior performance across industrial, semantic, and medical benchmarks.

D. Anomaly Localization

In Figure 3, we show CDAD anomaly localization results, with qualitative results in Figures 4, 5, and 6. CDAD excels in both Image-AUROC and Pixel-AUROC. On APTOS (Figure 4), retinal lesions are precisely localized, supporting medical diagnosis. On MPDD, MVTec-AD, and MVTec-LOCO (Figure 5), CDAD accurately segments industrial defects, confirming strong localization on industrial datasets. On Aircraft-FGVC (Figure 6), CDAD generates interpretable maps for semantic anomaly detection, where anomalous regions indicate the cause of abnormality.

E. Ablation Study

As depicted in Table II, we ablate three components on MPDD and MVTec-LOCO: focal loss, adaptor, and Conv module (Conv). Setting 1 is the baseline without any. Settings 2-4 add individual modules (focal loss+adaptor, focal loss+Conv, adaptor+Conv). Setting 5 incorporates all three, achieving the best performance and validating each module's contribution.

TABLE II

Ablation study results. Performance under five settings.

	Focal loss	Adaptor	Conv	MPDD	MVTec-LOCO
setting 1				97.80	84.86
setting 2	✓	✓		97.90	84.54
setting 3	✓		✓	97.90	85.18
setting 4		✓	✓	98.05	85.38
setting 5	✓	✓	✓	98.10	85.46

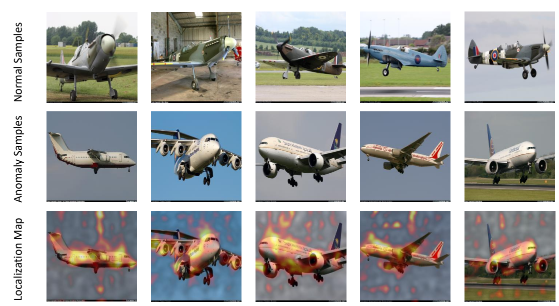

Fig. 6. **Semantic visualization.** CDAD on Aircraft-FGVC.

V. Conclusion

In this paper, we proposed CDAD, a unified model for cross-domain anomaly detection. It leverages a pre-trained feature extractor with adapted representations and a novel MFFD module to detect both local and global anomalies. A hybrid loss improves focus on hard samples. CDAD achieves excellent results across five datasets, demonstrating strong generalization and effectively bridging cross-domain gaps.

References

[1] M. Salehi, H. Mirzaei, D. Hendrycks, Y. Li, M. H. Rohban, and M. Sabokrou, "A unified survey on anomaly, novelty, open-set, and out-of-distribution detection: Solutions and future challenges," *arXiv preprint arXiv:2110.14051*, 2021.

[2] J. Guo, S. Lu, L. Jia *et al.*, "Recontrast: Domain-specific anomaly detection via contrastive reconstruction," *NIPS*, vol. 36, pp. 10 721–10 740, 2023.

[3] C. Huang, A. Jiang, J. Feng, Y. Zhang, X. Wang, and Y. Wang, "Adapting visual-language models for generalizable anomaly detection in medical images," in *CVPR*, 2024, pp. 11 375–11 385.

[4] M. J. Cohen and S. Avidan, "Transformaly-two (feature spaces) are better than one," in *CVPR*, 2022, pp. 4060–4069.

[5] H. Deng and X. Li, "Anomaly detection via reverse distillation from one-class embedding," in *CVPR*, 2022, pp. 9737–9746.

[6] H. Mirzaei, M. Salehi, S. Shahabi, E. Gavves, C. G. Snoek, M. Sabokrou, and M. H. Rohban, "Fake it until you make it: Towards accurate near-distribution novelty detection," in *ICLR*, 2022.

[7] T. Reiss and Y. Hoshen, "Mean-shifted contrastive loss for anomaly detection," in *AAAI*, vol. 37, no. 2, 2023, pp. 2155–2162.

[8] M. Salehi, N. Sadjadi, S. Baselizadeh, M. H. Rohban, and H. R. Rabiee, "Multiresolution knowledge distillation for anomaly detection," in *CVPR*, 2021, pp. 14 902–14 912.

[9] B. Bozorgtabar and D. Mahapatra, "Attention-conditioned augmentations for self-supervised anomaly detection and localization," in *AAAI*, vol. 37, no. 12, 2023, pp. 14 720–14 728.

[10] Z. Liu, Y. Zhou, Y. Xu, and Z. Wang, "Simplenet: A simple network for image anomaly detection and localization," in *CVPR*, 2023, pp. 20 402–20 411.

[11] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *CVPR*, 2022, pp. 14 318–14 328.

[12] Z. You, L. Cui, Y. Shen, K. Yang, X. Lu, Y. Zheng, and X. Le, "A unified model for multi-class anomaly detection," *NIPS*, vol. 35, pp. 4571–4584, 2022.

[13] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection," in *CVPR*, 2019, pp. 9592–9600.

[14] S. Maji, E. Rahtu, J. Kannala, M. Blaschko, and A. Vedaldi, "Fine-grained visual classification of aircraft," *arXiv preprint arXiv:1306.5151*, 2013.

[15] L. P. Sträter, M. Salehi, E. Gavves, C. G. Snoek, and Y. M. Asano, "Generalad: Anomaly detection across domains by attending to distorted features," in *ECCV*. Springer, 2024, pp. 448–465.

[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *ICCV*, 2017, pp. 2980–2988.

[17] S. Jezek, M. Jonak, R. Burget, P. Dvorak, and M. Skotak, "Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions," in *2021 13th International congress on ultra modern telecommunications and control systems and workshops (ICUMT)*. IEEE, 2021, pp. 66–71.

[18] P. Bergmann, K. Batzner, M. Fauser, D. Sattlegger, and C. Steger, "Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization," *International Journal of Computer Vision*, vol. 130, no. 4, pp. 947–969, 2022.

[19] Asia Pacific Tele-Ophthalmology Society, "APTOS 2019 blindness detection," 2019. [Online]. Available: https://www.kaggle.com/competitions/aptos2019-blindness-detection

[20] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "Cutpaste: Self-supervised learning for anomaly detection and localization," in *CVPR*, 2021, pp. 9664–9674.

[21] V. Zavrtanik, M. Kristan, and D. Skočaj, "Draem-a discriminatively trained reconstruction embedding for surface anomaly detection," in *ICCV*, 2021, pp. 8330–8339.

[22] X. Zhang, M. Xu, and X. Zhou, "Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection," in *CVPR*, 2024, pp. 16 699–16 708.

[23] T. Cao, J. Zhu, and G. Pang, "Anomaly detection under distribution shift," in *ICCV*, 2023, pp. 6511–6523.

[24] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "Padim: a patch distribution modeling framework for anomaly detection and localization," in *International conference on pattern recognition*. Springer, 2021, pp. 475–489.

[25] D. Gudovskiy, S. Ishizaka, and K. Kozuka, "Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows," in *WACV*, 2022, pp. 98–107.

[26] X. Xie, Y. Huang, W. Ning, D. Wu, Z. Li, and H. Yang, "Rdad: A reconstructive and discriminative anomaly detection model based on transformer," *International Journal of Intelligent Systems*, vol. 37, no. 11, pp. 8928–8946, 2022.

[27] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.

[28] T. Reiss, N. Cohen, L. Bergman, and Y. Hoshen, "Panda: Adapting pretrained features for anomaly detection and segmentation," in *CVPR*, 2021, pp. 2806–2814.

[29] D. Hendrycks, M. Mazeika, S. Kadavath, and D. Song, "Using self-supervised learning can improve model robustness and uncertainty," *NIPS*, vol. 32, 2019.

[30] J. Tack, S. Mo, J. Jeong, and J. Shin, "Csi: Novelty detection via contrastive learning on distributionally shifted instances," *NIPS*, vol. 33, pp. 11 839–11 852, 2020.

[31] J. Ma, W. Xie, H. Ye, D. Li, and L. Fang, "Aligning and prompting anything for zero-shot generalized anomaly detection," in *AAAI*, vol. 39, no. 6, 2025, pp. 5964–5972.

[32] Z. Gu, B. Zhu, G. Zhu, Y. Chen, M. Tang, and J. Wang, "Univad: A training-free unified model for few-shot visual anomaly detection," in *CVPR*, 2025, pp. 15 194–15 203.

[33] C. Ding, G. Pang, and C. Shen, "Catching both gray and black swans: Open-set supervised anomaly detection," in *CVPR*, 2022, pp. 7388–7398.

[34] J. Zhu, C. Ding, Y. Tian, and G. Pang, "Anomaly heterogeneity learning for open-set supervised anomaly detection," in *CVPR*, 2024, pp. 17 616–17 626.

[35] M. Caron, H. Touvron, I. Misra, H. Jégou, J. Mairal, P. Bojanowski, and A. Joulin, "Emerging properties in self-supervised vision transformers," in *ICCV*, 2021, pp. 9650–9660.

[36] M. Oquab, T. Darcet, T. Moutakanni, H. Vo, M. Szafraniec, V. Khalidov, P. Fernandez, D. Haziza, F. Massa, A. El-Nouby *et al.*, "Dinov2: Learning robust visual features without supervision," *arXiv preprint arXiv:2304.07193*, 2023.

[37] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, pp. 211–252, 2015.

[38] M. Salehi, E. Gavves, C. G. Snoek, and Y. M. Asano, "Time does tell: Self-supervised time-tuning of dense image representations," in *ICCV*, 2023, pp. 16 536–16 547.

[39] A. Ziegler and Y. M. Asano, "Self-supervised learning of object parts for semantic segmentation," in *CVPR*, 2022, pp. 14 502–14 511.

[40] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *NIPS*, vol. 30, 2017.

[41] X. Ma, X. Dai, Y. Bai, Y. Wang, and Y. Fu, "Rewrite the stars," in *CVPR*, 2024, pp. 5694–5703.

[42] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," *arXiv preprint arXiv:1711.05101*, 2017.