Preprint – IJCAI 2025: This is the accepted version made available for conference attendees.
Do not cite. The final version will appear in the IJCAI 2025 proceedings.

HyperTrans: Efficient Hypergraph-Driven Cross-Domain Pattern Transfer in Image Anomaly Detection

Tengyu Zhang[1] , **Deyu Zeng**[2*] , **Baoqiang Li**[2,3] , **Wei Wang**[3*] , **Wei Liu**[4] , **Zongze Wu**[2,1,3*]
[1]School of Software Engineering, Xi'an Jiaotong University,
[2]Shenzhen University,
[3]Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ),
[4]Wuhan University
{deyuzeng, zzwu}@szu.edu.cn, wangwei@gml.ac.cn

Abstract

Anomaly detection plays a pivotal role in industrial quality assurance processes, with cross-domain problems, exemplified by the model upgrade from RGB to 3D, being prevalent in real-world scenarios yet remaining systematically underexplored. To address the severe challenges posed by the extreme lack of datasets in target domain, we retain the knowledge from source models and explore a novel solution for anomaly detection through cross-domain learning, introducing HyperTrans. Targeting few-shot scenarios, HyperTrans centers around hypergraphs to model the relationship of the limited patch features and employs a perturbation-rectification-scoring architecture. The domain perturbation module injects and adapts channel-level statistical perturbations, mitigating style shifts during domain transfer. Subsequently, a residual hypergraph restoration module utilizes a cross-domain hypergraph to capture higher-order correlations in patches and align them across domains. Ultimately, with feature patterns exhibiting reduced domain shifts, an inter-domain scoring module aggregates similarity information between patches and normal patterns within the multi-domain sub-hypergraphs to make an integrated decision, generating multi-level anomaly predictions. Extensive experiments demonstrate that HyperTrans offers significant advantages in anomaly classification and anomaly segmentation tasks, outperforming state-of-the-art non-cross-domain methods in image-wise ROCAUC by 13%, 12%, and 15% in 1-shot, 2-shot, and 5-shot settings on MVTec3D AD.

1 Introduction

Image anomaly detection aims to identify abnormal images and segment anomalous subregions, playing a vital role in industrial quality control [Defard *et al.*, 2021; Roth *et al.*, 2022]. Traditional methods have predominantly relied on the RGB modality, achieving significant progress [Zuo *et al.*, 2024b; Li *et al.*, 2024]. However, advancements in technology have facilitated the adoption of image acquisition devices with multiple modalities, introducing pervasive cross-domain [Torralba and Efros, 2011] challenges. Among these, the shift to 3D is the most representative, as 3D sensors provide unique advantages for detecting depth-related defects [Zavrtanik *et al.*, 2024]. This transition has increased demand for model upgrades. Various approaches have been proposed from a non-cross-domain perspective, including point cloud-based methods [Wang *et al.*, 2023] and depth map-based methods [Chen *et al.*, 2023; Liu *et al.*, 2023], applied in both single-modal and multi-modal scenarios.

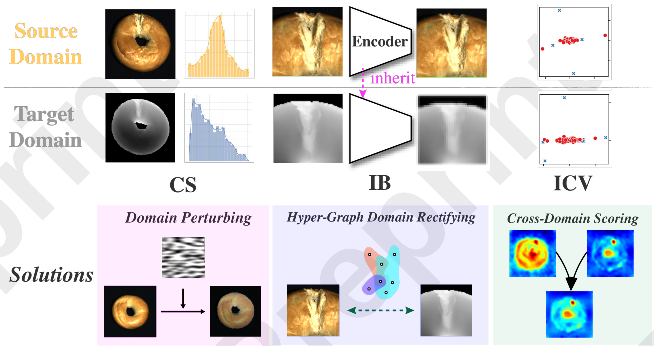

Figure 1: Challenges of cross-domain anomaly detection (Up) and our solutions (Down). A representative example is illustrated: source domain with RGB images and target domain with depth images. CS is short for 'Change of Style'. IB stands for 'Indefinable Boundaries'. ICV means 'Inter-class Variance'.

However, these non-cross-domain methods encounter inherent challenges. Multi-modal approaches, which rely on the collaborative input of both source and target data, impose high demands on acquisition devices and modality alignment. Single-modal methods, which retrain the source-specific model, tend to discard valuable information embedded in the original source model-information, optimized with extensive validated data. Furthermore, all existing approaches encounter substantial challenges due to the limited availability of target datasets, which are characterized by extreme scarcity and insufficient labeled samples. This limitation hinders the effective training of large-scale target models, particularly during the initial phases when acquiring adequate

---

[*] Co-Correpsonding Authors.

labeled data remains a resource-intensive endeavor.

To address these challenges and preserve the effective supervisory signals in the source model, we introduce and explore the feasibility of cross-domain (CD) adaptation [Wilson and Cook, 2020; Zhang *et al.*, 2022] in industrial anomaly detection. This approach transfers and fine-tunes the source model for other domains. To further mitigate the issue of limited target domain data, we emphasize few-shot learning (FS) scenarios [Huang *et al.*, 2022; Xie *et al.*, 2023a]. Specifically, we conduct the first dedicated study on the problem of Cross-Domain Few-Shot Image Anomaly Detection (CD-FSAD), which transfers knowledge from the source domain and leverages few-shot samples from the target domain, providing an innovative solution to anomaly detection.

Cross-domain adaptation in defect detection poses significant challenges. As illustrated in Figure 1, images captured by different sensors exhibit significant disparities. Achieving domain alignment is impeded by three key factors: 1) changes in image style (CS), which hinder feature extraction; 2) blurred boundaries (IB), which degrade anomaly segmentation accuracy; and 3) inter-class variance (ICV) [Fu *et al.*, 2025], which compromises defect classification precision. To systematically address these challenges, we propose an innovative perturbation-rectification-scoring framework named HyperTrans, which maximizes the exploitation of source-domain information by hierarchically aligning and scoring patches across domains, achieving cross-domain generality and defect-detection specificity.

Specifically, a domain perturbation module injects channel-level perturbations to simulate new domain styles, followed by a lightweight adapter to align these features with the original domain, improving generalization. Furthermore, limited samples exhibit inherent relational information, as different patches from both domains may share similar properties, such as reflecting translation and rotation invariance of the same object attributes. To capture this higher-order correlations and enhance boundary delineation, we construct cross-domain hypergraphs [Hamilton *et al.*, 2017; Gao *et al.*, 2020; Kipf and Welling, 2022; Feng *et al.*, 2023] that link patches globally across samples and domains. A residual hypergraph convolutional network (ResHGNN) further integrates and refines these latent relationships, with pseudo-features designated for target domain features, enabling effective cross-domain feature alignment.

While improving target domain adaptability, the previous training strategy also brings the source and target domain patches closer, thereby enhancing the utility of the source domain bank. Consequently, we incorporate the source domain bank into inference and propose a cross-domain scoring module leveraging subhypergraphs to assign anomaly scores and segmentation maps. By synthesizing differential outcomes between normal patterns across domains and query features, the module offers a unified solution to ICV and enables more precise anomaly classification and segmentation.

Our main contributions are listed as follows:

1) We pioneer the application of domain adaptation to image anomaly detection and propose an innovative perturbation-rectification-scoring framework tailored for cross-domain few-shot scenarios, facilitating efficient knowledge transfer between domains with minimal labeled data.
2) We design a compact hypergraph cross-domain anomaly detection module. Centered on hypergraph, three key functions are implemented: a residual hypergraph convolutional network to model cross-domain higher-order correlations, a pseudo-feature generation method to guide the alignment and a cross-domain scoring method leveraging cross-domain normal patch patterns to facilitate anomaly detection.
3) Comprehensive experiments validate the superiority of the proposed model, significantly outperforming other state-of-the-art non-cross-domain methods in few-shot anomaly detection tasks and establishing a benchmark for cross-domain few-shot anomaly detection.

2 Related Works

2.1 Image Anomaly Detection

Anomaly detection identifies rare data points that indicate defects in industrial production [Defard *et al.*, 2021; Roth *et al.*, 2022]. Similar to surface anomaly detection [Zavrtanik *et al.*, 2022], both aim to recognize unexpected patterns in data or images. Methods for anomaly detection can be broadly classified into two categories: contrastive reasoning based methods and meta-learning based methods. Among the contrastive methods, image-level contrastive reasoning focuses on comparing entire images, such as in SimpleNet [Liu *et al.*, 2023], which learns simpler representations for anomaly detection through image-level reconstruction. Feature-level contrastive reasoning compares features extracted from images, and includes methods like PatchCore [Roth *et al.*, 2022], which performs anomaly detection by comparing feature patches. Instead, DSR [Zavrtanik *et al.*, 2022], which trains a general VQ-VAE [Van Den Oord *et al.*, 2017] and a specialized decoder to output anomaly maps, can be categorized as a meta-learning-based method. Similarly, AST [Rudolph *et al.*, 2023] is also a meta-learning based method, as it employs a student-teacher architecture and utilizes knowledge distillation for anomaly detection. However, neither method can solve the cross-domain problem, as its model cannot adapt well to the diverse styles of features on the target domain.

2.2 Few-shot and Cross-domain Learning

In few-shot defect detection, many methods have been explored. For example, CLIP-FSAC [Zuo *et al.*, 2024a] leverages the high generalization ability of the CLIP model and uses an alternately trained adapters to enable few-shot defect classification, demonstrating that adapters are effective for transferring the model to new domains. Additionally, graph-based methods have shown promising results by exploiting the correlation information between samples. The graph structure can effectively reduce the reliance on the number of samples by leveraging information from a few samples [Xie *et al.*, 2023b]. In this work, we generalize this idea to hypergraph and extend it to model cross-domain feature patterns. Although cross-domain approaches have not been exclusively explored in defect detection, many general cross-domain frameworks [Muandet *et al.*, 2013; Liu *et al.*, 2024;

Fu *et al.*, 2025] have emerged in recent years and can serve as paradigms. For instance, PDA [Bai *et al.*, 2024] also inherits the high generalization capability of the CLIP model and utilizes cross-domain memory banks with a cross-attention enhancement mechanism, addressing the domain adaptation problem and more relevantly, demonstrating the effectiveness of memory banks in cross-domain tasks. DR-Adapter [Su *et al.*, 2024] achieves top results in cross-domain few-shot segmentation through perturbations and a novel cyclic perturbation correction model. However, it is a one-to-one correction mode and lacks targeted exploration of relationships in small sample settings. By leveraging these approaches and uncovering high-order patch correlations, we introduce cross-domain few-shot learning to anomaly detection via the HyperTrans framework, which is tailored for defect detection and holds broader cross-domain applicability.

3 Methodology

3.1 Problem Definition

A typical defect detection task includes anomaly classification (AC) and anomaly segmentation (AD) [Li *et al.*, 2024]. The dataset is divided into a normal sample set N and an anomaly sample set A. The goal is to train a model $f(\cdot)$ in an unsupervised manner, using only normal samples from dataset $x \in N$, to produce output consisting of a binary classification result for defect detection (denoted as a) and a binary semantic segmentation result for defect localization (denoted as S), where the classification result a is more primary. The whole process can be expressed as:

$$a, S = f(x), \begin{cases} x \in N & \text{, if training} \\ x \in N \cup A & \text{, if testing} \end{cases}, \tag{1}$$

Cross-domain few-shot anomaly detection aims at generalizing and transferring the applicability from the source domain to the target domain, where only a few normal samples are available for training. Formally, given a model $f(\cdot)$ trained with data from the source domain $x \in N_s$ and few-shot target domain normal samples $N_t^{\text{few-shot}}$, the task is to align the model to mitigate the issue of domain discrepancy between $N_s \cup A_s$ and $N_t \cup A_t$. The training process is conducted on the source domain followed by fine-tuning on the target domain, and the inference process is performed on the target domain, which can be denoted as:

$$a, S = f(x), \begin{cases} x \in N_s \cup N_t^{\text{few-shot}} & \text{, if training} \\ x \in N_t \cup A_t & \text{, if testing} \end{cases}. \tag{2}$$

3.2 Hyper-graph Cross-domain Few-shot Anomaly Detection Framework

Overview

To accomplish CD-FSAD, we propose a novel framework called HyperTrans, with its architecture illustrated in Figure 2. The training process is firstly conducted on the source-domain, where the domain adapter is optimized with perturbed patches and a source-domain feature bank is generated. Correspondingly the target feature bank is generated with few-shot target samples. The two banks are united to construct a cross-domain hyper-graph, representing the higher-order correlations between patches. Subsequently, a pseudo-feature is assigned to each feature through the hyper-graph, and the distances between them are reduced by the hyper-graph residual network to correct the cross-domain feature shift. During the inference process, the the source and target domain subhypergraphs are separately contracted along hyperedges to compute the anomaly score of patches, and the vertice-level results are weighted and integrated to form the image-wise anomaly score and pixel-wise anomaly map.

Domain Perturbation

Previous studies [Zhou *et al.*, 2021; Su *et al.*, 2024] have shown that perturbing channel statistics can effectively simulate different domain styles, thereby enhancing the domain adaptability. Building on this insight, we apply this technique to simulate diverse image patches extracted by the backbone.

For clarity, we denote the feature extractor symbolically as $\mathcal{B}(\cdot)$, which derives patch features $F = \mathcal{B}(N_s)$ from the source samples N_s. Given the feature set $F \in \mathbb{R}^{n \times c \times h \times w}$, the mean $\text{avg}(\cdot)$ and variance $\text{std}(\cdot)$ are first calculated to represent feature channel statistics:

$$\begin{aligned} \text{avg}(F) &= \frac{1}{hw} \sum_{i=1}^{h} \sum_{j=1}^{w} F, \\ \text{std}(F) &= \sqrt{\frac{1}{hw} \sum_{i=1}^{h} \sum_{j=1}^{w} (F - \text{avg}(F))^2}, \end{aligned} \tag{3}$$

where n is the number of patches, c is the number of channels, h and w represent the height and weight of the patches. In defect detection, changes are typically localized and abrupt. Therefore, instead of using Gaussian noise, we design an alternative mechanism based on Perlin noise [Perlin, 1985] to better approximate smooth local variations in spatial structures and perturb the channel statistics (mean and variance) of feature maps. Specifically, two perturbation factors, P_{avg} and P_{std}, are generated using Perlin noise, with dimensions matching the feature channels. These factors are applied to adjust the original mean and standard deviation as follows:

$$\begin{aligned} \text{avg}_p(F) &= \text{avg}(F) + P_{\text{avg}} \cdot \text{avg}(F), \\ \text{std}_p(F) &= \text{std}(F) + P_{\text{std}} \cdot \text{std}(F). \end{aligned} \tag{4}$$

The perturbed statistics are then integrated into the Adaptive Instance Normalization (AdaIN) formula [Huang and Belongie, 2017] to compute the modified feature map:

$$\omega(F) = \text{std}_p(F) \cdot \frac{F - \text{avg}(F)}{\text{std}(F)} + \text{avg}_p(F). \tag{5}$$

To simplify, this can be rewritten as:

$$\omega(F) = (1 + P_{\text{std}}) \cdot F + (P_{\text{avg}} - P_{\text{std}}) \cdot \text{avg}(F). \tag{6}$$

Such perturbations are accommodated by a learnable adapter of the deep layers of the backbone. To bring the simulated features F_p closer to the original features F_o, a domain perturbation loss with Mean Squared Error (MSE) is developed as the objective function, formally expressed as:

$$\mathcal{L}_{dp} = \frac{1}{n} \|F - \omega(F)\|^2, \tag{7}$$

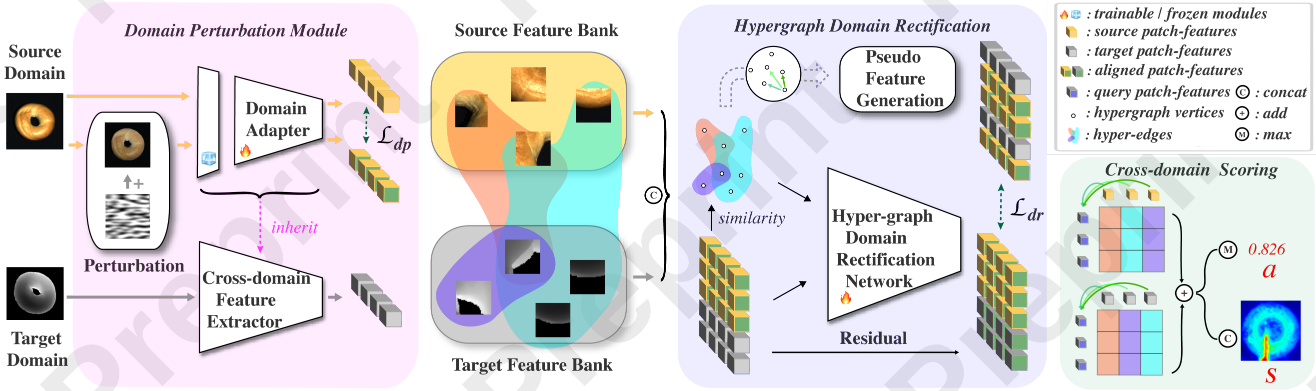

Figure 2: The architecture of HyperTrans.

where $\|\cdot\|^2$ denotes calculating the 2-norm of the matrix. n is the number of elements in the patch.

This approach dynamically adjusts the perturbation of the mean and variance using Perlin noise, leveraging its smooth, spatially coherent properties. Unlike direct noise addition, this method seamlessly integrates the perturbations with the feature map's channel statistics, resulting in spatially consistent and locally adaptive feature variations. Subsequently, through $\mathcal{L}_{dp}$, the adaptability of the original backbone to different domain styles is augmented.

Hyper-graph Domain Rectification

To mitigate the domain discrepancy of patch features across different domains, we designed a cross-domain hypergraph feature correction module (HDR) centered on hypergraphs. Samples from the source and target domains represent different manifestations of the same object, with intra-domain and inter-domain patches exhibiting correlations—such as describing the same attributes of an object. Hypergraph is utilized to bridge such high-order correlations. Unlike edges in simple graphs, hyperedges in hypergraphs naturally simulate the attribute features shared among multiple vertices.

Each core patch in the banks from different domains is treated as a vertex. If let F_s denote the set of patches from the source domain and F_t denote the patches from the target domain, the cross-domain hypergraph can be represented as $\mathcal{G} = \{F_s \cup F_t, \mathcal{E}, W\}$, where $\mathcal{E}$ is the hyperedge set. $W \in \mathbb{R}^{|\mathcal{E}|\times|\mathcal{E}|}$ is a diagonal matrix, where $W_{i,i}$ denotes the weight of the i-th hyperedge.

The hyperedges are constructed based on the nodes with the reciprocal of Euclidean distance being the similarity metric. A hyperedge is created for each node, connecting it to other nodes based on similarity. For a source feature bank containing $|F_s|$ patches and a target bank with $|F_t|$ patches, a total of $|F_s| + |F_t|$ hyperedges are constructed in this manner.

For computational convenience, the hypergraph is represented in matrix form as $H \in [0,1]^{|F_s \cup F_t|\times|\mathcal{E}|}$ where each row represents the vertex and each column represents a hyperedge. The value of $H_{i,j} \in [0,1]$ indicates the correlation from i-th vertex to the j-th hyperedge. To leverage the high-order relational information embedded in the hypergraph to guide feature rectification, a hypergraph convolution [Gao *et al.*, 2022] (HGNNConv) is employed. It is based on hypergraph convolution layers $c(\cdot)$, defined as follows:

$$c(F) = \sigma\left(D_v^{-\frac{1}{2}} H W D_e^{-1} H^\top D_v^{-\frac{1}{2}} F \Theta\right), \tag{8}$$

where D_v and D_e are the diagonal degree matrices for vertex and hyperedge, respectively. $\Theta \in \mathbb{R}$ is the trainable parameter for the HGNNConv layer. By applying HGNNConv on the hypergraph constructed from cross-domain patches, the generated structure-aware embeddings can capture latent collaborative information from both the source and target domains.

The final hypergraph network is constructed with two layers of HGNNConv. Additionally, a residual structure is equipped to preserve the original features and reduce the compromise of feature correction on the original information. The overall output of the Res-HGNN is expressed as:

$$\mathcal{R}(F) = \alpha c_1(\text{dropout}(\text{sigmoid}(c_2(F)))) + (1-\alpha)F, \tag{9}$$

where α is a hyperparameter to balance the corrected and original features. The dropout rate is set to 0.5.

The pseudo feature generation is then developed to assign a prototype to each patch by leveraging the correlation information within the hypergraph. Specifically, for each node in the hypergraph, the feature of the most correlated node within its corresponding hyperedges is selected as the pseudo feature. In matrix form, the generation process can be denoted as:

$$\mathcal{S}(F) = g(F, \text{ max\_index}(H^c)), \tag{10}$$

where $g(a, b)$ represents extracting elements from matrix a based on index list b and max_index$(\cdot)$ outputs the index of the largest component for each row. In H^c, non-cross-domain connections are set to 0.

To align the output features with the pseudo features and the original ones, Mean Squared Error is adopted:

$$\mathcal{L}_{dr} = \frac{1}{n}\|\mathcal{R}(F) - \mathcal{S}(F)\|^2 + \frac{1}{n}\|\mathcal{R}(F) - F\|^2. \tag{11}$$

Cross-domain Scoring

The feature bank from the source domain stores extensive information about normal samples. Relying on the feature bank for inference solely on target domain samples is insufficient, as it neglects the valuable information from source domain. Initially, due to domain shift, the differences between the source domain's feature bank and the original normal features of the target domain are significant, greatly impacting the effectiveness of defect scoring. However, the preceding domain perturbation and rectification modules align the target domain features generated by the model with the source domain features, mitigating the effects of domain shift. This alignment enhances the utility of the source domain bank for comparative scoring in the target domain. To further capitalize on this, we propose a cross-domain scoring module.

For each image I from the target domain requiring inference, we denote its patch feature set as $F = \mathcal{R}(\mathcal{B}(I))$, from which f is a specific patch feature. In the first step, the patches to be inferred are incorporated alongside the source and target domain banks to build the hypergraph H. To separately leverage the similarity information from both the source and target domains, the hypergraph matrix H is split. The source domain subhypergraph $H^s \in [0,1]^{|F|\times|F_s|}$ preserves the correlations between target features to be inferred and source feature bank, while the target domain subhypergraph $H^t \in [0,1]^{|F|\times|F_t|}$ preserves correlations of query target features and target feature bank. Subsequently, a defect score is assigned to each patch leveraging the subhypergraph from the source domain H^s, and the maximum score among all patches is taken as the image-level anomaly score:

$$f^* = \arg\max_{f\in F} \max(H^s(f)), \tag{12}$$

$$\psi(H^s, f) = \max(H^s(f^*)), \tag{13}$$

where $\max(\cdot)$ returns the max element of a vector and $H^s(f)$ is the row with f being the vertex in the subhypergraph matrix in the source domain H^s. The pixel-wise anomaly map is obtained by rearranging patch anomaly scores based on their respective spatial positions:

$$\phi(H^s, f) = \{\max(H^s(f)) | f \in F\}. \tag{14}$$

The procedure is also adopted to assign anomaly score and segmentation map with the target domain subhypergraph:

$$f^* = \arg\max_{f\in F} \max(H^t(f)), \tag{15}$$

$$\psi(H^t, f) = \max(H^t(f^*)), \tag{16}$$

$$\phi(H^t, f) = \{\max(H^t(f)) | f \in F\}. \tag{17}$$

The final image-wise anomaly score a and pixel-wise anomaly map S are obtained by combining the results of multiple domains with a parameter β to balance the impact of information from source and target domain:

$$a = \beta\psi(H_s, f) + (1-\beta)\psi(H_t, f), \tag{18}$$

$$S = \beta\phi(H_s, f) + (1-\beta)\phi(H_t, f). \tag{19}$$

To match the original input resolution, the final anomaly map is upsampled using bilinear interpolation. Additionally, a Gaussian function with a kernel width of $\sigma = 4$ is applied to smooth the results.

3.3 Model Training

Benefiting from the above observation, we propose a two-stage learning strategy to train the whole model. The first stage will employ the domain perturbation loss Formula (7) to optimize the domain adapter in DP. Then, the model switch into the second stage with the domain rectification loss, i.e., Formula (11), to optimize the parameters in Res-HGNN. The overall procedure is illustrated in Algorithm 1.

Algorithm 1 Optimization algorithm for HyperTrans

Input: Backbone $\mathcal{B}(\cdot)$, domain perturbing module $\omega(\cdot)$, normal samples from source domain N_s, few-shot normal samples from target domain N_t, original parameters in domain adapter Θ, initialized parameters of Res-HGNN $\{\Theta_1, \Theta_2\}$
Parameters: *epochsP*, *epochsR*, ...
Output: Memory banks $\{\mathcal{M}_s, \mathcal{M}_t\}$, optimized parameters $\{\Theta, \Theta_1, \Theta_2\}$

```
while training do
    for $e = 1 \to epochsP$ do
        $F \leftarrow \mathcal{B}(N_s),\ F_p \leftarrow \omega(F)$
        $\Theta \xleftarrow{optim} \mathcal{L}_{dp}(F, F_p)$
    end for
    while constructing feature banks do
        $F_s \leftarrow \mathcal{B}(N_s), F_t \leftarrow \mathcal{B}(N_t)$
        $\mathcal{M}_s \xleftarrow{store} F_s,\ \mathcal{M}_t \xleftarrow{store} F_t$
    end while
    for $e = 1 \to epochsR$ do
        $F \leftarrow \mathcal{B}(N_t) \cup \mathcal{M}_s,$
        $F_r \leftarrow \mathcal{R}(\mathcal{B}(F)),\ F_{pse} \leftarrow \mathcal{S}(F)$
        $\{\Theta_1, \Theta_2\} \xleftarrow{optim} \mathcal{L}_{dr}(F_r, F_{pse})$
    end for
end while
```

4 Experiments

4.1 Experimental Settings

Dataset: The MVTec-3D AD dataset [Bergmann *et al.*, 2021] is a standard 3D anomaly detection dataset, on which our experiments are performed. It consists of 10 categories, a total of 2656 training samples, and 1137 testing samples. The 3D scans were acquired by a high-resolution industrial 3D sensor using structured light. Due to the scarcity of defective samples, only normal samples are contained in the training set of MVTec-3D AD, while the testing set contains both normal and defective samples. The following experiments are carried out on it, with RGB data as the source domain and the XYZ data preprocessed into depth images to be the target domain.

Implementation Details: We implement HyperTrans using PyTorch2.4.0 and conduct evaluations on an NVIDIA A100. The ViT_base_patch8_224_dino [Caron *et al.*, 2021] serves as the backbone. The domain perturbation module is integrated into layers 3 to 10 of the backbone, employing Perlin noise with a mean of 0 and a standard deviation of 0.75 for domain perturbations.[1]

Evaluation Metrics: Models are evaluated with ROCAUC in image and pixel orientation (Receiver operating characteristic - Area under the curve), which is a metric to evaluate the performance of a classifier by calculating the area under the ROC curve at various thresholds. The image-wise ROCAUC evaluates the performance based on the classification of entire images as defective or non-defective, while the pixel-wise ROCAUC assesses the ability to correctly classify each individual pixel as part of a defect or not. Initial values of other parameters are listed in Table 2.

---

[1] https://github.com/raRn0y/HyperTrans

Task	Type	Method	Backbone	bagel	cable_gland	carrot	cookie	dowel	foam	peach	potato	rope	tire	AVG[1]
Img[2]	Target 1-shot	Depth iNet [Horwitz and Hoshen, 2023]	ResNet	0.486	0.587	0.615	0.814	0.598	0.589	0.460	0.584	0.619	0.625	0.598
		SIFT [Horwitz and Hoshen, 2023]	SIFT	0.502	0.476	0.503	0.651	0.688 ①	0.574	0.553	0.536	0.580	0.541	0.560
		AST [Rudolph *et al.*, 2023]	EfficientNet	0.642	0.598	0.693	0.821②	0.581	0.532	0.654	0.595	0.911②	0.531	0.656
		3DSR [Zavrtanik *et al.*, 2024]	VQ-VAE	0.665	0.613①	0.562	0.610	0.599	0.601②	0.531	0.603②	0.651	0.519	0.595
		M3DM [Wang *et al.*, 2023]	ViT	0.754②	0.549	0.718②	0.582	0.64②	0.791①	0.772②	0.576	0.801	0.68①	0.686②
		HyperTrans	ViT	0.816①	0.606②	0.957①	0.955①	0.463	0.569	0.926①	0.926①	0.921①	0.64②	0.778①
	Target 2-shot	Depth iNet [Horwitz and Hoshen, 2023]	ResNet	0.522	0.496	0.639	0.822	0.599	0.529	0.475	0.663②	0.662	0.584②	0.599
		SIFT [Horwitz and Hoshen, 2023]	SIFT	0.562	0.420	0.543	0.579	0.613②	0.634②	0.567	0.506	0.769	0.530	0.572
		AST [Rudolph *et al.*, 2023]	EfficientNet	0.722	0.602	0.713	0.84②	0.548	0.505	0.676	0.570	0.909①	0.522	0.654
		3DSR [Zavrtanik *et al.*, 2024]	VQ-VAE	0.565	0.623①	0.575	0.531	0.568	0.551	0.567	0.521	0.624	0.567	0.569
		M3DM [Wang *et al.*, 2023]	ViT	0.909②	0.607②	0.762②	0.626	0.683①	0.708①	0.809②	0.601	0.827	0.571	0.71②
		HyperTrans	ViT	0.93①	0.581	0.965①	0.958①	0.551	0.489	0.962①	0.99①	0.909①	0.646①	0.798①
	Target 5-shot	Depth iNet [Horwitz and Hoshen, 2023]	ResNet	0.592	0.504	0.675	0.852②	0.675②	0.599	0.474	0.683	0.708	0.526	0.629
		SIFT [Horwitz and Hoshen, 2023]	SIFT	0.568	0.371	0.601	0.777	0.614	0.591	0.530	0.689②	0.607	0.528	0.588
		AST [Rudolph *et al.*, 2023]	EfficientNet	0.739	0.519	0.639	0.764	0.455	0.556	0.601	0.658	0.863②	0.445	0.624
		3DSR [Zavrtanik *et al.*, 2024]	VQ-VAE	0.603	0.600	0.578	0.576	0.613	0.627②	0.599	0.600	0.672	0.602②	0.607
		M3DM [Wang *et al.*, 2023]	ViT	0.921①	0.631①	0.781②	0.585	0.831①	0.704①	0.863②	0.579	0.801	0.512	0.721②
		HyperTrans	ViT	0.92②	0.631①	0.998①	0.981①	0.604	0.621	0.954①	0.996①	0.904①	0.697①	0.831①
Pix[3]	Target 1-shot	Depth iNet [Horwitz and Hoshen, 2023]	ResNet	0.940	0.911	0.966	0.963②	0.938	0.744	0.951	0.958	0.962	0.597	0.893
		SIFT [Horwitz and Hoshen, 2023]	SIFT	0.969②	0.923②	0.988②	0.931	0.942	0.876②	0.974②	0.993②	0.971②	0.923①	0.949②
		AST [Rudolph *et al.*, 2023]	EfficientNet	0.929	0.516	0.925	0.628	0.543	0.556	0.684	0.853	0.910	0.590	0.713
		3DSR [Zavrtanik *et al.*, 2024]	VQ-VAE	0.441	0.498	0.868	0.705	0.747	0.573	0.657	0.906	0.362	0.567	0.632
		M3DM [Wang *et al.*, 2023]	ViT	0.954	0.859	0.965	0.959	0.956①	0.826	0.959	0.913	0.967	0.247	0.861
		HyperTrans	ViT	0.984①	0.935①	0.997①	0.976①	0.954②	0.895①	0.995①	0.998①	0.992①	0.821②	0.955①
	Target 2-shot	Depth iNet [Horwitz and Hoshen, 2023]	ResNet	0.944	0.918	0.966	0.967②	0.913	0.752	0.957	0.960	0.966	0.497	0.884
		SIFT [Horwitz and Hoshen, 2023]	SIFT	0.973②	0.927②	0.99②	0.932	0.943	0.897②	0.972②	0.993②	0.978②	0.941①	0.955②
		AST [Rudolph *et al.*, 2023]	EfficientNet	0.940	0.507	0.925	0.639	0.539	0.564	0.678	0.863	0.929	0.522	0.711
		3DSR [Zavrtanik *et al.*, 2024]	VQ-VAE	0.544	0.810	0.884	0.692	0.748	0.612	0.601	0.870	0.411	0.611	0.678
		M3DM [Wang *et al.*, 2023]	ViT	0.960	0.840	0.965	0.962	0.956②	0.804	0.967	0.910	0.968	0.629	0.896
		HyperTrans	ViT	0.989①	0.937①	0.998①	0.984①	0.967①	0.901①	0.996①	0.999①	0.995①	0.834②	0.96①
	Target 5-shot	Depth iNet [Horwitz and Hoshen, 2023]	ResNet	0.946	0.884	0.967	0.967②	0.944	0.747	0.959	0.958	0.971	0.561	0.890
		SIFT [Horwitz and Hoshen, 2023]	SIFT	0.974②	0.925②	0.993②	0.940	0.956②	0.89②	0.977②	0.995②	0.976②	0.931①	0.956②
		AST [Rudolph *et al.*, 2023]	EfficientNet	0.958	0.519	0.867	0.614	0.485	0.594	0.601	0.944	0.864	0.445	0.689
		3DSR [Zavrtanik *et al.*, 2024]	VQ-VAE	0.619	0.833	0.675	0.692	0.500	0.633	0.390	0.569	0.420	0.717	0.605
		M3DM [Wang *et al.*, 2023]	ViT	0.962	0.770	0.968	0.960	0.959①	0.775	0.970	0.912	0.973	0.556	0.881
		HyperTrans	ViT	0.992①	0.935①	0.997①	0.98①	0.955	0.895①	0.996①	0.999①	0.993①	0.9②	0.964①

Table 1: Comparison results of anomaly classification and anomaly segmentation on MVTec-3D AD dataset. The first place and second place are marked. [1]'AVG' is the average of all categories. [2]'Img' stands for anomaly classification with image-wise ROCAUC being the evaluation metric. [3]'Pix' is short for anomaly segmentation with pixel-wise ROCAUC being the evaluation metric.

α	β	$epochsP$	$epochsR$
0.3	0.6	3	100

Table 2: Parameter settings of HyperTrans in our experiments.

4.2 Comparision Experiments

Statistics: To verify the few-shot general anomaly detection performance, we compare HyperTrans with advanced depth methods. The image-wise and pixel-wise ROCAUCs are reported in Table 1. It can be concluded that source models generally have limited domain adaptability. When handling tasks with domain shift, most methods suffer from great accuracy loss, most of which come from the model design targeting source characteristics and the deep fitting of network to source data. With only a few depth samples (1, 2 and 5 samples for each classes) given, HyperTrans achieves the SOTA depth-domain performance in the task of cross-domain anomaly classification and segmentation. The advantage is more significant on the primary metric of image-wise ROCAUC with an improvement of 13%, 12% and 15% on 1, 2 and 5 shot settings compared with the best non-cross-domain methods. Tracing the image-wise results, HyperTrans improves the distinguishability on anomaly samples and alleviates the issue of inter-class variance. From the pixel-wise results, it enhances feature clarity and, through cross-domain joint inference-making, mitigates the problem of IB.

Visualizations: Figure 3 shows intuitive examples of predicted segmentation maps. HyperTrans achieves the best segmentation, effectively addressing the issue of IB.

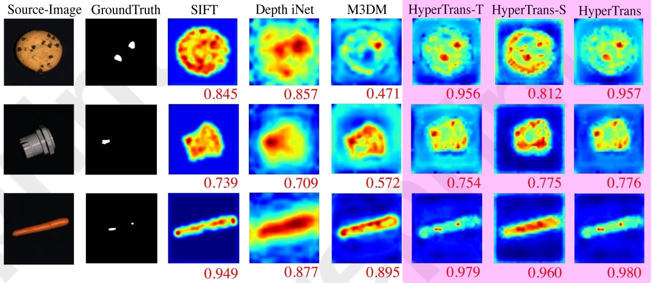

Figure 3: Qualitative visualization of segmentation masks conducted on MVTec3D AD under 1-shot setting. It compares HyperTrans with baselines and illustrates the effectiveness of joint reasoning using source (-S) and target (-T) memory banks. AU-PRO scores in the lower right reflect the localization accuracy of anomalies.

4.3 Parameter Analysis

To evaluate the impact of α (residual on hypergraph cross-domain feature alignment), β (the influence of the source and target domains during inference), and the cross-interaction between feature alignment and inference, we conducted a series of parameter analysis experiments on the two parameters. The results are shown in Figure 4. As α increases, the cross-domain enhancement component increases, while the residual component decreases. From Figure 4 (Left), it can be seen that due to the fitting of HDR module on the source do-

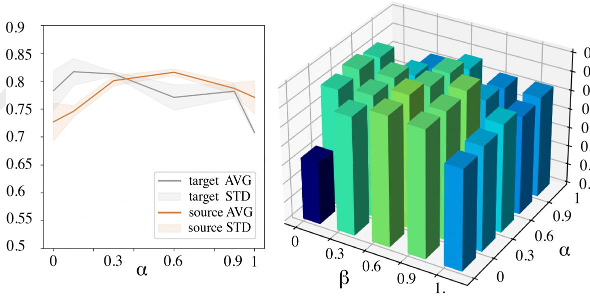

Figure 4: Parameter analysis of HyperTrans. The impact of α on the results for source and target domains (Left) and the cross impact of α and β (Right). The results are the average image-wise ROCAUC among all categories on MVTec-3D AD.

DP	HDR	CDS	I-ROCAUC
✗	✗	✗	0.813 ± 0.203
✓	✗	✗	0.817 ± 0.257
✗	✓	✗	0.826 ± 0.217
✗	✗	✓	0.817 ± 0.233
✓	✓	✓	0.831 ± 0.158

Table 3: Ablation study results of the proposed modules. 'DP' is short for domain perturbation. 'HDR' represents for hypergraph domain rectification. 'CDS' represents cross-domain scoring. 'I-ROCAUC' is the average image-wise ROCAUC of all classes on MVTec3D AD in terms of mean ± standard deviation.

main, the results relying on source domain patterns show an increase in fluctuations, while the results on the target domain gradually decrease. Therefore it's necessary to introduce β to balance the weight of source and target domain for inference and study the cross impact of α and β. From Figure 4 (Right), it can be concluded that HyperTrans achieves the best average ROCAUC on the test set when $\alpha = 0.3$, $\beta = 0.6$ and the introduction of both components is essential.

4.4 Ablation Studies

Impact of Each Modules: To measure the impact of each losses on the results and the effectiveness of each component in the framework, we conduct an ablation analysis with results shown in Table 3. We observe a consistent performance improvement with the introduction of each proposed modules, and the combination of them improves the averaged results by 0.018. This improvement is attributed to the effectiveness of cross-domain alignment, which not only mitigates the issues highlighted in Figure 1 but also brings the source and target domain patches closer. This proximity enhances the utility of the source domain bank, further improving the inference performance of the CDS mechanism.

Impact of Domain Perturbation: The DP module aims to simulate samples from various domains with channel statistic noise injection, guiding the model to adapt to non-existent domain styles in advance. t-SNE [Van der Maaten and Hinton, 2008] in Figure 5 shows that the perturbed features, after undergoing correction through this module, become closer to the original ones, demonstrating its effectiveness.

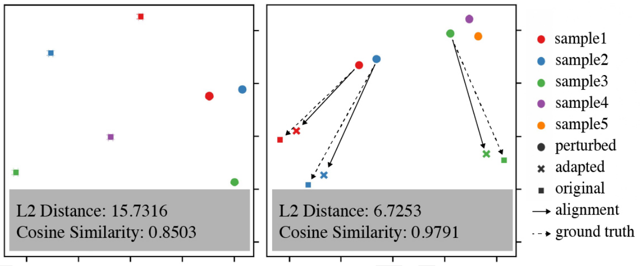

Figure 5: t-SNE on the features before (Left) and after (Right) Domain Perturbing. All features are generated from bagel on MVTec3D AD. 'L2 Distance' is the average Euclidean Distance between the adapted features and original features. 'Cosine Similarity' computes the average cosine similarity between the two.

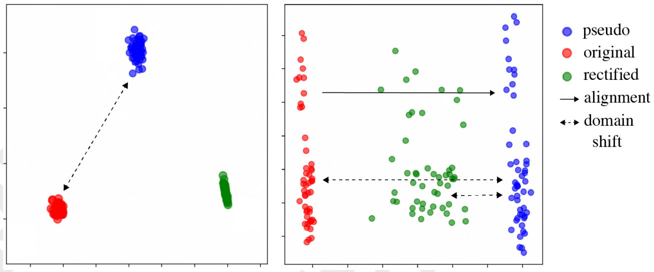

Figure 6: PCA on the features before Hypergraph Domain Rectification (Left) and features after Hypergraph Domain Rectification (Right). All features are generated from bagel on MVTec3D AD.

Impact of Hypergraph Domain Rectification: The HDR module provides a global correction across domains. We visualize its impact with PCA [Jolliffe and Cadima, 2016] in Figure 6. Comparing the distributions of the original patches with those of the HDR-enhanced patches, the representation spaces of the source and target samples become closer and the features are more aligned. This reduces the domain shift and enhances the utility of source bank during inference.

5 Summary

In this paper, we pioneer the exploration of cross-domain learning in image anomaly detection and introduce CDFSAD, a task designed to enable the effective transfer of knowledge for anomaly detection in few-shot learning scenarios. A novel perturbation-rectification-scoring solution named HyperTrans is then developed. The domain perturbation module simulates target domain styles using statistical Perlin noise, while the domain rectification module employs hypergraphs to capture higher-order cross-domain feature correlations, integrating residual hypergraph convolution and pseudo-feature generation for alignment. During inference, a cross-domain scoring method jointly leveraging multi-domain subhypergraph similarity information for multi-level anomaly reasoning. Extensive experiments validate that HyperTrans achieves SOTA performance, significantly outperforming non-cross-domain methods in classification and segmentation tasks, and providing a scalable, efficient and universal solution for image anomaly detection.

Acknowledgements

This work was supported in part by the Grants of Guangdong Major Project of Basic and Applied Basic Research (2023B0303000009) and the National Major Scientific Instruments and Equipments Development Project of National Natural Science Foundation of China under Grant (62327808) and the Shenzhen University 2035 Program for Excellent Research (00000224), and the National Natural Science Foundation of China Project (62403326).

References

[Bai *et al.*, 2024] Shuanghao Bai, Min Zhang, Wanqi Zhou, Siteng Huang, Zhirong Luan, Donglin Wang, and Badong Chen. Prompt-based distribution alignment for unsupervised domain adaptation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 729–737, 2024.

[Bergmann *et al.*, 2021] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. *arXiv preprint arXiv:2112.09045*, 2021.

[Caron *et al.*, 2021] Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9650–9660, 2021.

[Chen *et al.*, 2023] Ruitao Chen, Guoyang Xie, Jiaqi Liu, Jinbao Wang, Ziqi Luo, Jinfan Wang, and Feng Zheng. Easynet: An easy network for 3d industrial anomaly detection. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7038–7046, 2023.

[Defard *et al.*, 2021] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International Conference on Pattern Recognition*, pages 475–489. Springer, 2021.

[Feng *et al.*, 2023] Yifan Feng, Shuyi Ji, Yu-Shen Liu, Shaoyi Du, Qionghai Dai, and Yue Gao. Hypergraph-based multi-modal representation for open-set 3d object retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[Fu *et al.*, 2025] Yuqian Fu, Yu Wang, Yixuan Pan, Lian Huai, Xingyu Qiu, Zeyu Shangguan, Tong Liu, Yanwei Fu, Luc Van Gool, and Xingqun Jiang. Cross-domain few-shot object detection via enhanced open-set object detector. In *European Conference on Computer Vision*, pages 247–264. Springer, 2025.

[Gao *et al.*, 2020] Yue Gao, Zizhao Zhang, Haojie Lin, Xibin Zhao, Shaoyi Du, and Changqing Zou. Hypergraph learning: Methods and practices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(5):2548–2566, 2020.

[Gao *et al.*, 2022] Yue Gao, Yifan Feng, Shuyi Ji, and Rongrong Ji. Hgnn+: General hypergraph neural networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):3181–3199, 2022.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Advances in neural information processing systems*, 30, 2017.

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

[Huang and Belongie, 2017] Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pages 1501–1510, 2017.

[Huang *et al.*, 2022] Chaoqin Huang, Haoyan Guan, Aofan Jiang, Ya Zhang, Michael Spratling, and Yan-Feng Wang. Registration based few-shot anomaly detection. In *European Conference on Computer Vision*, pages 303–319. Springer, 2022.

[Jolliffe and Cadima, 2016] Ian T Jolliffe and Jorge Cadima. Principal component analysis: a review and recent developments. *Philosophical transactions of the royal society A: Mathematical, Physical and Engineering Sciences*, 374(2065):20150202, 2016.

[Kipf and Welling, 2022] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2022.

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

[Liu *et al.*, 2023] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20402–20411, 2023.

[Liu *et al.*, 2024] Yingnan Liu, Yingtian Zou, Rui Qiao, Fusheng Liu, Mong Li Lee, and Wynne Hsu. Cross-domain feature augmentation for domain generalization. In *International Joint Conference on Artificial Intelligence*, 2024.

[Muandet *et al.*, 2013] Krikamol Muandet, David Balduzzi, and Bernhard Schölkopf. Domain generalization via invariant feature representation. In *International conference on machine learning*, pages 10–18. PMLR, 2013.

[Perlin, 1985] Ken Perlin. An image synthesizer. *ACM Siggraph Computer Graphics*, 19(3):287–296, 1985.

[Roth *et al.*, 2022] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter

Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328, 2022.

[Rudolph *et al.*, 2023] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2592–2602, 2023.

[Su *et al.*, 2024] Jiapeng Su, Qi Fan, Wenjie Pei, Guangming Lu, and Fanglin Chen. Domain-rectifying adapter for cross-domain few-shot segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24036–24045, 2024.

[Torralba and Efros, 2011] Antonio Torralba and Alexei A Efros. Unbiased look at dataset bias. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1521–1528. IEEE, 2011.

[Van Den Oord *et al.*, 2017] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

[Van der Maaten and Hinton, 2008] Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

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

[Wilson and Cook, 2020] Garrett Wilson and Diane J Cook. A survey of unsupervised deep domain adaptation. *ACM Transactions on Intelligent Systems and Technology*, 11(5):1–46, 2020.

[Xie *et al.*, 2023a] Guoyang Xie, Jingbao Wang, Jiaqi Liu, Feng Zheng, and Yaochu Jin. Pushing the limits of fewshot anomaly detection in industry vision: Graphcore. In *International Conference on Learning Representations*, 2023.

[Xie *et al.*, 2023b] Guoyang Xie, Jingbao Wang, Jiaqi Liu, Feng Zheng, and Yaochu Jin. Pushing the limits of fewshot anomaly detection in industry vision: Graphcore. In *International Conference on Learning Representations*, 2023.

[Zavrtanik *et al.*, 2022] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr–a dual subspace re-projection network for surface anomaly detection. In *European conference on computer vision*, pages 539–554. Springer, 2022.

[Zavrtanik *et al.*, 2024] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Cheating depth: Enhancing 3d surface anomaly detection via depth simulation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2164–2172, 2024.

[Zhang *et al.*, 2022] Min Zhang, Siteng Huang, and Donglin Wang. Domain generalized few-shot image classification via meta regularization network. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3748–3752. IEEE, 2022.

[Zhou *et al.*, 2021] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. In *International Conference on Learning Representations*, 2021.

[Zuo *et al.*, 2024a] Zuo Zuo, Yao Wu, Baoqiang Li, Jiahao Dong, You Zhou, Lei Zhou, Yanyun Qu, and Zongze Wu. Clip-fsac: Boosting clip for few-shot anomaly classification with synthetic anomalies. In *International Joint Conference on Artificial Intelligence*, 2024.

[Zuo *et al.*, 2024b] Zuo Zuo, Zongze Wu, Badong Chen, and Xiaopin Zhong. A reconstruction-based feature adaptation for anomaly detection with self-supervised multi-scale aggregation. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 5840–5844. IEEE, 2024.