Registration based Few-Shot Anomaly Detection

Chaoqin Huang[1,3,4], Haoyan Guan[2], Aofan Jiang[1], Ya Zhang[1,3] ✉,
Michael Spratling[2], and Yan-Feng Wang[1,3] ✉

[1] Cooperative Medianet Innovation Center, Shanghai Jiao Tong University
{huangchaoqin, stillunnamed, ya_zhang, wangyanfeng}@sjtu.edu.cn
[2] King's College London
{haoyan.guan, michael.spratling}@kcl.ac.uk
[3] Shanghai Artificial Intelligence Laboratory
[4] National University of Singapore

Abstract. This paper considers few-shot anomaly detection (FSAD), a practical yet under-studied setting for anomaly detection (AD), where only a limited number of normal images are provided for each category at training. So far, existing FSAD studies follow the one-model-per-category learning paradigm used for standard AD, and the inter-category commonality has not been explored. Inspired by how humans detect anomalies, *i.e.,* comparing an image in question to normal images, we here leverage registration, an image alignment task that is inherently generalizable across categories, as the proxy task, to train a category-agnostic anomaly detection model. During testing, the anomalies are identified by comparing the registered features of the test image and its corresponding support (normal) images. As far as we know, this is the first FSAD method that trains a single generalizable model and requires no re-training or parameter fine-tuning for new categories. Experimental results have shown that the proposed method outperforms the state-of-the-art FSAD methods by 3%-8% in AUC on the MVTec and MPDD benchmarks. Source code is available at: https://github.com/MediaBrain-SJTU/RegAD

Keywords: Anomaly Detection, Few-Shot Learning, Registration

1 Introduction

Anomaly detection (AD), with a wide range of applications such as defect detection [24], medical diagnosis [44], and autonomous driving [10], has received quite some attention in the computer vision community over the last decades. With the ambiguous definition of "anomaly", *i.e.,* samples that do not conform to the "normal", it is impossible to train with an exhaustive set of anomalous samples. As a result, recent studies on anomaly detection have largely been devoted to unsupervised learning, *i.e.,* learning with only the "normal" samples. Through modeling the normal distribution with one-class classification [35,30,43], reconstruction [47,13,39,18], or self-supervised learning tasks [12,42,33,45], many AD

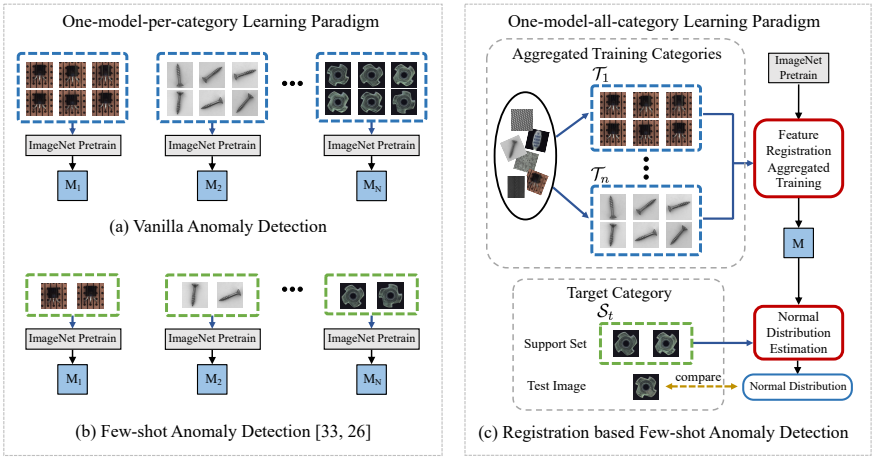

Fig. 1. Different from (a) vanilla AD, and (b) existing FSAD methods under the one-model-per-category learning paradigm, the proposed method (c) leverages feature registration as a category-agnostic approach for FSAD, under the one-model-all-category learning paradigm. Trained with aggregated data of multiple categories, the model is directly applicable to novel categories without any parameter fine-tuning, with the only need to estimate the normal feature distribution given the corresponding support set.

methods detect anomalies by identifying samples with different distributions than the model.

Most existing AD methods have focused on training a dedicated model for each category (Fig. 1 (a)). However, in real-world scenarios such as defect detection, given hundreds of industrial products to handle, it is not cost-effective to collect a large training set for each product, not to mention the need for many time-sensitive applications. A couple of studies [36,29] have recently explored a special, yet practical, setting of AD, *i.e.,* few-shot anomaly detection (FSAD), where only a limited number of normal images are provided for each category at training (Fig. 1 (b)). The few-shot learning of anomaly detection has been approached with strategies to reduce the demand on training samples, such as radical data augmentation with multiple transformations [36] or a lighter estimator for the normal distribution estimation [29]. *However, such approaches still follow the one-model-per-category learning paradigm and fail to leverage the inter-category commonality.*

This paper aims to explore a new paradigm for FSAD, by learning a common model shared among multiple categories and also generalizable to novel categories, and inspired by how human beings detect anomalies. In fact, when a human is asked to search for the anomaly in an image, a simple strategy one may adopt is to compare the sample to a normal one to find the difference. As long as one knows how to compare two images, the actual semantics of the images does not matter anymore. To achieve such a human-like comparison process, we resort to registration, a process of transforming different images into one co-

ordinate system in order to better enable comparison [4,46,25]. Registration is particularly suitable for FSAD, as *registration is expected to be category-agnostic and thus generalizable across categories, allowing the model to be adaptable to novel categories without the necessity of parameter fine-tuning.*

Fig. 1 (c) provides an overview of the proposed Registration based few-shot Anomaly Detection (RegAD) framework. To train a category-agnostic anomaly detection model, we leverage registration, a task that is inherently generalizable across categories, as the proxy task. A Siamese network [5] with three spatial transformer network [19] blocks is employed as the registration network (see Fig. 2). For better robustness, instead of registering the images pixel-by-pixel as typical registration methods [25], here we propose a feature-level registration loss by maximizing the cosine similarity of features from the same category, which may be deemed as a relaxed version of the pixel-wise registration loss. Normal images from different categories are used together to aggregately train the model, with two images from the same category randomly selected as a training pair. Such aggregated training procedure is adopted so as to enable the trained registration model to be category-agnostic. At test time, a support set of a few normal samples is provided for the target category, together with each test sample. It is straightforward to identify anomalies by comparing the registered features of the test image and the corresponding support (normal) images. Given the support set, the normal distribution of registered features for the target category is estimated with a statistical-based distribution estimator [8]. Test samples that are out of the statistical normal distribution are considered anomalies. In this way, the model quickly adapts to novel categories by simply estimating its normal feature distribution without any parameter fine-tuning.

To validate the effectiveness of RegAD, we experiment with two challenging benchmark datasets for industrial defect detection, MVTec AD [2] and MPDD [20]. Our experimental results have shown that RegAD outperforms the state-of-the-art FSAD methods [36,29], achieving improvements of 5.1%, 6.9%, and 8.0% in AUC on MVTec, and improvements of 3.2%, 5.0%, and 3.4% in AUC on MPDD, for 2-shot, 4-shot, and 8-shot scenarios, respectively.

The main contributions of the paper are summarised as follows:

- We introduce feature registration as a category-agnostic approach for few-shot anomaly detection (FSAD). To our best of knowledge, it is the first FSAD method that trains a single generalizable model and requires no re-training or parameter fine-tuning for new categories.
- Extensive experiments on recent benchmark datasets have shown that the proposed RegAD outperforms the state-of-the-art FSAD methods on both the anomaly detection and anomaly localization tasks.

2 Related Work

2.1 Anomaly Detection

AD is a task where training datasets contain only normal data. To better estimate the normal distributions, one-class classification based approaches tend

to depict the normal data directly with statistical approaches [9,35,26,30]. Self-supervised based approaches are trained using only normal data, and then make inferences by assuming that anomalous data performs differently. In this domain, reconstruction [40,34,47,32,1,13,39,17] is the most popular self-supervision. Some approaches [12,42,33] introduce other self-supervisions, *e.g.*, [12] applies dozens of image geometric transforms for transformation classification; [42] proposes a restoration framework for attribute restoration. Recent AD methods usually use feature embeddings extracted from a pre-trained deep neural network. Feature embedding is mostly used as an input for a traditional machine learning algorithm or statistical metrics such as the Mahalanobis distance [8]. The network used as a feature extractor can be trained from scratch [43], while several methods [21,8,45,28,14] have also achieved state-of-the-art results using models pre-trained on the ImageNet dataset [31]. This paper differs from these previous works by focusing on FSAD, where only a few normal images are available.

2.2 Few-shot Learning

Few-shot learning (FSL) aims to adapt to novel classes with a few annotated examples. Representative FSL methods can be categorized into metric learning, generation, and optimization. Metric learning approaches [37,38,15] learn to calculate a feature space that classifies an unseen sample based on its nearest example category. Generation methods [22,41,6] enhance the novel class performance by generating its images or features. Optimization methods [27,11] learn commonalities among different categories and explore efficient optimization strategies for novel classes based on these commonalities. In this paper, the proposed method predicts 'normal' or 'anomaly' for a new category. In contrast to previous work on FSL, both training data and support set only have positive (normal) examples without any negative (anomaly) samples.

2.3 Few-shot Anomaly Detection

FSAD aims to indicate anomalies with only a few normal samples as the support images for target categories. TDG [36] proposes a hierarchical generative model that captures the multi-scale patch distribution of each support image. They use multiple image transformations and optimize discriminators to distinguish between real and fake patches, as well as between different transformations applied to the patches. The anomaly score was obtained by aggregating the patch-based votes of the correct transformations. DiffNet [29] leverages the descriptiveness of features extracted by convolutional neural networks to estimate their density using a normalizing flow, which is a tool well-suited to estimate distributions from a few support samples. Metaformer [39] can be applied to the FSAD, although an additional large-scale dataset, MSRA10K [7], should be used during its entire meta-training procedure (beyond parameter pre-training), together with additional pixel-level annotations. In this paper, we design registration based FSAD to learn the category-agnostic feature registration, enabling the model to detect anomalies in new categories given a few normal images without fine-tuning.

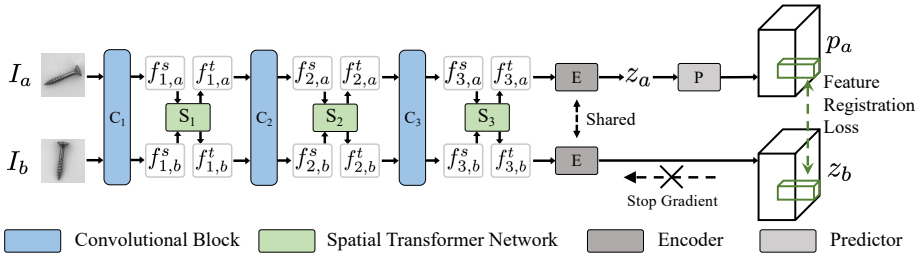

Fig. 2. The model architecture of the proposed RegAD. Given paired images from the same category, features are extracted by three convolutional residual blocks each followed by a spatial transformer network. A Siamese network acts as the feature encoder, supervised by a registration loss for feature similarity maximization.

3 Problem Setting

We first formally define the problem setting for the proposed few-shot anomaly detection. Given a training set consisting of only normal samples of n categories, *i.e.*, $\mathcal{T}_{train} = \bigcup_{i=1}^{n} \mathcal{T}_i$, where the subset $\mathcal{T}_i$ consists of normal samples from the category c_i, $(i = 1, 2, \cdots, n)$, we want to train a category-agnostic anomaly detection model. At test time, given a normal or anomalous image from a target category c_t $(t \notin \{1, 2, \cdots, n\})$ and its associated support set $\mathcal{S}_t$ consisting of k normal samples from the target category c_t, the trained category-agnostic anomaly detection model should predict whether the image is anomalous or not.

For FSAD, we attempt to detect anomalies from test samples of unseen/novel categories using only a few normal images as the support set. The key challenges lie in: (i) $\mathcal{T}_{train}$ has only access to normal samples from multiple known categories (*e.g.*, different objects or textures), without any image-level or pixel-level annotations, (ii) the test data is from an unseen/novel category, and (iii) only a few normal samples from the target category c_t are available, making it hard to estimate the normal distribution of the target category c_t.

4 Method

Motivated by how humans detect anomalies, the feature registration is used as a generalization paradigm for FSAD. During the training procedure, we leverage an anomaly-free feature registration network to learn category-agnostic feature registration. During testing, given the support set of a few normal images, the normal distribution of registered features for the target category is estimated with a statistical-based distribution estimator. Test samples that are out of the learned statistical normal distribution are considered anomalies.

4.1 Feature Registration Network

Given a pair of images I_a and I_b randomly selected from a same category in the training set $\mathcal{T}_{train}$, a ResNet-type convolutional network [16] is leveraged as the

feature extractor. Specifically, as shown in Fig. 2, the first three convolutional residual blocks of ResNet, C_1, C_2, and C_3, are adopted, and the last convolution block in ResNet's original design is discarded, in order to ensure that final features still retain spatial information. A spatial transformer network (STN) [19] is inserted into each block as a feature transformation module, so as to enable the model to learn feature registration flexibly, inspired by [45]. Specifically, a transformation function S_i $(i = 1, 2, 3)$ is applied on an input feature f_i^s:

$$\begin{pmatrix} x_i^t \\ y_i^t \end{pmatrix} = S_i(f_i^s) = A_i \begin{pmatrix} x_i^s \\ y_i^s \\ 1 \end{pmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix} \begin{pmatrix} x_i^s \\ y_i^s \\ 1 \end{pmatrix}, \tag{1}$$

where (x_i^t, y_i^t) are the target coordinates of output feature f_i^t, (x_i^s, y_i^s) are the same points in the source coordinates of input feature f_i^s and A_i is the affine transformation matrix. The module S_i is used to learn the mappings from features of convolutional block C_i with the same tiny architecture as used in [19].

Given paired extracted features $f_{3,a}^t$ and $f_{3,b}^t$ as the final transformation outputs, we design the feature encoder as a Siamese network [3]. A Siamese network is a parameter-sharing neural network applied on multiple inputs. To avoid the collapsing problem when optimized without negative pairs, inspired by SimSiam [5], features are processed by the same encoder network E followed by a prediction head P applied on one branch. A stop-gradient operation is applied on the other branch, as shown in Fig. 2, which is critical to prevent such collapsing solutions. Denote $p_a \triangleq P(E(f_{3,a}))$ and $z_b \triangleq E(f_{3,b})$, a negative cosine similarity loss is applied:

$$\mathcal{D}(p_a, z_b) = -\frac{p_a}{||p_a||_2} \cdot \frac{z_b}{||z_b||_2}, \tag{2}$$

where $||\cdot||_2$ is a L_2 norm. Instead of registering the images pixel-by-pixel, here we use a feature-level registration loss which may be deemed as a relaxed version of the pixel-wise registration constraints for better robustness. Finally, following SimSiam [5], a symmetrized feature registration loss is defined as:

$$\mathcal{L} = \frac{1}{2}(\mathcal{D}(p_a, z_b) + \mathcal{D}(p_b, z_a)). \tag{3}$$

Discussion. Features from the proposed method retain relatively complete spatial information, since we adopt the first three convolutional blocks of ResNet as the backbone without global average pooling, followed by a convolutional encoder and predictor architecture, but not the MLP architecture in SimSiam [5]. Thus Eq. (3) should be computed by averaging cosine similarity scores at every spatial pixel. Features containing spatial information are beneficial for the AD task, which needs to provide anomaly score maps as prediction results. Different from SimSiam [5], which defines the inputs as two augmentations of one image and maximizes their similarity to enhance the model representation, the proposed feature registration leverages two different images as inputs and maximizes the similarity between the features to learn the registration.

4.2 Normal Distribution Estimation

To perform testing, it is assumed that the feature registration ability can generalize to the target category, and the learned feature registration model is applied to the support set $\mathcal{S}_t$ for the target category without parameter fine-tuning. Multiple data augmentations are applied to the support images, consistent with [36]. As the two branches of the Siamese network are exactly the same, only one branch feature is used for the normal distribution estimation. After achieving the registered features, a statistical-based estimator [8] is used to estimate the normal distribution of target category features, which uses multivariate Gaussian distributions to get a probabilistic representation of the normal class. Suppose an image is divided into a grid of $(i, j) \in [1, W] \times [1, H]$ positions where $W \times H$ is the resolution of features used to estimate the normal distribution. At each patch position (i, j), let $F_{ij} = \{f_{ij}^k, k \in [1, N]\}$ be the registered features from N augmented support images. f_{ij} is the aggregated features at patch position (i, j), achieved by concatenating the three STN outputs at the corresponding position with upsampling operations to match their sizes. By the assumption that F_{ij} is generated by $\mathcal{N}(\mu_{ij}, \Sigma_{ij})$, the sample covariance is:

$$\Sigma_{ij} = \frac{1}{N-1} \sum_{k=1}^{N} \left(f_{ij}^k - \mu_{ij}\right) \left(f_{ij}^k - \mu_{ij}\right)^{\mathrm{T}} + \epsilon I, \tag{4}$$

where μ_{ij} is the sample mean of F_{ij}, and the regularization term ϵI makes the sample covariance matrix full rank and invertible. Finally, each possible patch position is associated with a multivariate Gaussian distribution.

Discussion. Data augmentations are widely adopted in AD, and especially in FSAD, including TDG [36] and DiffNet [29]. However, most methods simply apply the data augmentations on both the support and test images without any exploration of the impact. In this paper, we emphasize that data augmentation plays a very important role in expanding the support set, which is beneficial for the normal distribution estimation. Specifically, we adopt augmentations including rotation, translation, flipping, and graying for each image in the support set $\mathcal{S}_t$. Other augmentations like mixup and cutpaste are not considered since they seem more suitable for simulating anomalies [21]. We conduct the possible combinations of all these augmentations for each sample in the support set, which jointly combine into a larger support set. We conduct the normal distribution estimation on such an augmented support set. We study the impacts of different augmentations in the supplementary material.

4.3 Inference

During inference, test samples that are out of the normal distribution are considered anomalies. For each test image in $\mathcal{T}_{test}$, we use the Mahalanobis distance $\mathcal{M}(f_{ij})$ to give an anomaly score to the patch in position (i, j), where

$$\mathcal{M}(f_{ij}) = \sqrt{(f_{ij} - \mu_{ij})^T \Sigma_{ij}^{-1} (f_{ij} - \mu_{ij})}. \tag{5}$$

The matrix of Mahalanobis distances $\mathcal{M} = (\mathcal{M}(f_{ij}))_{1\leqslant i\leqslant W, 1\leqslant j\leqslant H}$ forms an anomaly map. Three inverse affine transformations corresponding to the three STN modules are applied to this anomaly map to get the final anomaly score map $\mathcal{M}_{final}$ aligned with the original image. High scores in this map indicate the anomalous areas. The final anomaly score of the entire image is the maximum of anomaly map $\mathcal{M}_{final}$. Compared with [36,29], RegAD cancels the data augmentation of the test images which reduces the inference computational costs.

5 Experiments

5.1 Experimental Setups

Datasets. We experiment on two challenging real-world benchmark datasets for AD [2,20], which are both related to industrial defect detection.

- **MVTec** [2]: MVTec comprises 15 categories with 3629 images for training and validation and 1725 images for testing. The training set contains only of normal images without defects. The test set contains both images with various kinds of defects (anomaly) and defect-free images (normal). On average five per category, 73 different defect types are given. All images are in the resolution range between 700×700 and 1024×1024 pixels. Pixel-wise ground truth labels for each defective image region are provided.
- **MPDD** [20]: MPDD is a newly proposed dataset focused specifically on defect detection during painted metal part fabrication, containing 6 classes of metal parts. Images are captured under the conditions of various spatial orientations, positions, and distances of multiple objects, concerning different light intensities and a non-homogeneous background.

For each dataset, we conduct experiments on two different experimental settings. (i) **Aggregated training** on multiple categories and then adapting to unseen categories, and (ii) **Individual training** only with the support set for each category.

Competing Methods. We consider two state-of-the-art FSAD approaches, TDG [36] and DiffNet [29]. These two methods both train models individually for each category (setting (ii)). Results are reproduced using the official source code. Considering that our method uses data from multiple categories, for fairness of comparison, we extend them to leverage the same amount of data (setting (i)). A pre-training procedure is added to these methods, where data from multiple categories are used to pre-train the transformation classifier for TDG or initialize the normalizing flow-based estimator for DiffNet. The corresponding methods are TDG+ and DiffNet+. We also evaluate RegAD under the individual training setting, and denote the corresponding method as RegAD-L. We compare with some state-of-the-art vanilla AD methods, such as GANomaly [1], ARNet [42], MKD [33], CutPaste [21], FYD [45], PaDiM [8], PatchCore [28] and CflowAD [14]. These methods use the whole normal dataset for their training, so they can be deemed as the upper bound on FSAD performance.

Table 1. Results of k-shot anomaly detection on the MVTec dataset, comparing with state-of-the-art methods. Results are listed as the average AUC in % of 10 runs and are marked individually for each category. A macro-average score over all categories is also reported in the last row. The best-performing method is in bold.

Category	k=2			k=4			k=8		
	TDG+ [36]	DiffNet+ [29]	RegAD (ours)	TDG+ [36]	DiffNet+ [29]	RegAD (ours)	TDG+ [36]	DiffNet+ [29]	RegAD (ours)
Bottle	69.3	99.3	**99.4**	69.6	99.3	**99.4**	70.3	99.4	**99.8**
Cable	68.3	**85.3**	65.1	70.3	**85.2**	76.1	74.7	**87.9**	80.6
Capsule	55.1	**73.0**	67.5	47.6	**80.3**	72.4	44.7	**78.6**	76.3
Carpet	66.2	78.4	**96.5**	68.7	78.6	**97.9**	78.2	78.5	**98.5**
Grid	83.8	62.1	**84.0**	86.2	60.5	**91.2**	87.6	78.5	**91.5**
Hazelnut	67.2	94.9	**96.0**	71.2	**95.8**	95.8	82.8	**97.9**	96.5
Leather	93.6	90.7	**99.4**	93.2	91.2	**100**	93.5	92.2	**100**
Metal Nut	67.1	61.9	**91.4**	69.2	67.3	**94.6**	68.7	67.6	**98.3**
Pill	69.2	**83.2**	81.3	64.7	**84.0**	80.8	67.9	**82.1**	80.6
Screw	**98.8**	73.4	52.5	**98.8**	72.5	56.6	**99.0**	75.0	63.4
Tile	86.3	**97.0**	94.3	87.2	**98.0**	95.5	87.4	**99.6**	97.4
Toothbrush	54.4	60.8	**86.6**	57.8	62.5	**90.9**	57.6	60.8	**98.5**
Transistor	55.9	61.8	**86.0**	67.7	62.2	**85.2**	71.5	63.3	**93.4**
Wood	98.4	98.1	**99.2**	98.3	96.4	**98.6**	98.4	**99.4**	**99.4**
Zipper	64.4	**89.2**	86.3	65.3	84.8	**88.5**	66.3	87.3	**94.0**
Average	73.2	80.6	**85.7**	74.4	81.3	**88.2**	76.6	83.2	**91.2**

Evaluation Protocols. We quantify the model performance using the area under the Receiver Operating Characteristic (ROC) curve metric (AUC), which is commonly adopted as the performance measurement for AD tasks. The image-level AUC and the pixel-level AUC are used for anomaly detection and anomaly localization respectively.

Model Configuration and Training Details. An ImageNet pre-trained ResNet-18 [16] is used as the backbone, followed by a convolutional-based encoder and predictor. To retain the spatial information, the encoder contains three 1×1 convolutional layers, while the predictor contains two 1×1 convolutional layers, without any pooling operation. We train models on 224×224 images on one NVIDIA GTX 3090. We update the parameters using momentum SGD with a learning rate of 0.0001 for 50 epochs, with a batch size of 32. A single cycle of cosine learning rate is used as the decay schedule.

5.2 Comparison with State-of-the-art Methods

Comparison with Few-Shot Anomaly Detection Methods. Experiments were conducted using the leave-one-out setting, *i.e.*, a target category was chosen to be tested, while other categories in the dataset are used for training. Table 1 and Table 2 show the comparison results on MVTec and MPDD, respectively, under the experimental setting (i). RegAD achieves an improvement of 5.1%, 6.9%, 8.0% in average AUC on MVTec, and an improvement of 3.2%, 5.0%, 3.4% in average AUC on MPDD, over DiffNet+ [29], with 2-shot, 4-shot, and

Table 2. Results of k-shot anomaly detection on the MPDD dataset, comparing with state-of-the-art methods. Results are listed as the average AUC in % of 10 runs and are marked individually for each category. A macro-average score over all categories is also reported in the last row. The best-performing method is in bold.

Category	k=2			k=4			k=8		
	TDG+ [36]	DiffNet+ [29]	RegAD (ours)	TDG+ [36]	DiffNet+ [29]	RegAD (ours)	TDG+ [36]	DiffNet+ [29]	RegAD (ours)
bracket black	46.4	56.7	**63.3**	48.8	59.9	**63.8**	51.0	**69.7**	67.3
bracket brown	54.9	**61.3**	59.4	57.5	64.2	**66.1**	65.4	66.3	**69.6**
bracket white	**64.0**	42.2	55.6	**65.4**	51.8	59.3	66.8	**69.1**	61.4
connector	53.1	54.1	**73.0**	55.8	54.8	**77.2**	62.9	54.5	**84.9**
metal plate	91.8	**96.8**	61.7	95.1	**98.2**	78.6	98.4	**98.8**	80.2
tubes	51.8	49.8	**67.1**	58.5	50.7	**67.5**	64.9	52.6	**67.9**
Average	60.3	60.2	**63.4**	63.5	63.3	**68.3**	68.2	68.5	**71.9**

Table 3. Results of anomaly detection on the MVTec and MPDD datasets under two different experimental settings (i) and (ii), comparing with state-of-the-art few-shot anomaly detection methods on $k = 2, 4, 8$. Results are listed as the macro-average AUC in % over all categories in each dataset of 10 runs. The best-performing method for each experimental setting is in bold.

Methods	ImageNet Pretrain	Aggregated Training	Time of Adaptation	MVTec			MPDD		
				k=2	k=4	k=8	k=2	k=4	k=8
TDG [36]	✓	✗	-	71.2	72.7	75.2	57.3	60.4	64.4
DiffNet [29]	✓	✗	-	80.5	80.8	82.9	**58.4**	**61.2**	**66.5**
RegAD-L (ours)	✓	✗	-	**81.5**	**84.9**	**87.4**	50.8	54.2	61.1
TDG+ [36]	✓	✓	1559.76s	73.2	74.4	76.6	60.3	63.5	68.2
DiffNet+ [29]	✓	✓	357.75s	80.6	81.3	83.2	60.2	63.3	68.5
RegAD (ours)	✓	✓	4.47s	**85.7**	**88.2**	**91.2**	**63.4**	**68.3**	**71.9**

8-shot scenarios, respectively. Also, with one-shot, RegAD achieves 82.4% and 57.8% AUC on MVtec and MPDD respectively.

RegAD is tested without any parameter fine-tuning, which may not guarantee the best performance for every category, while other baselines have unfair advantages in that they tune the parameters for each category. In 9 out of the 15 categories, RegAD outperforms all the other baselines. RegAD also achieves the least standard deviation (10.94) for the 15 categories when k=8, compared to TDG+ (15.20) and DiffNet+ (13.11), suggesting its better generalizability across different categories. Also, although using different training settings, for MVTec (k=8), RegAD achieves 91.2% AUC, with an ≈3% improvement compared with Metaformer [39] which uses an additional large-scale dataset, MSRA10K [7], during its entire training procedure.

Discussion. Adaptation time is important for real-world applications of FSAD. The procedures of fine-tuning for both TDG+ and DiffNet+ are time-consuming since they update the models for many epochs, while RegAD has the fastest adaptation speed since it is based on a statistical estimator which needs only one inference for each support image. In Table 3, we report the adaptation times for each method, by averaging the results for $k = 2, 4, 8$ on both the

Table 4. Results of anomaly detection and anomaly localization on the MVTec and MPDD datasets, comparing with state-of-the-art vanilla AD methods. Results are listed as AUC in % as the macro-average score over all categories in each dataset.

Methods	Data	ImageNet Pretrain	Backbone	MVTec		MPDD	
				image	pixel	image	pixel
RegAD (k=4)	4 images	✓	Res18	88.2	95.8	68.8	93.9
RegAD (k=8)	8 images	✓	Res18	91.2	96.7	71.9	95.1
RegAD (k=16)	16 images	✓	Res18	92.7	96.6	75.3	96.3
RegAD (k=32)	32 images	✓	Res18	94.6	96.9	76.8	96.3
GANomaly [1]	full data	✗	UNet	80.5	-	64.8	-
ARNet [42]	full data	✗	UNet	83.9	-	69.7	-
MKD [33]	full data	✓	Res18	87.7	90.7	-	-
CutPaste [21]	full data	✓	Res18	95.2	96.0	-	-
FYD [45]	full data	✓	Res18	97.3	97.4	-	-
PaDiM [8]	full data	✓	WRN50	97.9	97.5	74.8	96.7
PatchCore [28]	full data	✓	WRN50	99.1	98.1	82.1	95.7
CflowAD [14]	full data	✓	WRN50	98.3	98.6	86.1	97.7

MVTec and MPDD datasets. Compared with TDG+ (1559.76s) and DiffNet+ (357.75s), the proposed RegAD has the fastest adaptation speed (4.47s).

Table 3 also compares these methods under experimental setting (ii), where we train the models individually using the support images for each category. RegAD-L means RegAD with individual training on one category only. Assuming that features pre-trained by ImageNet are fully representative, we simply fine-tune features using limited support images. Thus, we conduct the fine-tuning procedures directly under an ImageNet pre-training backbone for all methods. All methods use the same ImageNet pre-training backbone to have a fair comparison. In this setting, RegAD-L outperforms both TDG and DiffNet on the MVTec dataset. DiffNet performs better than the proposed method on the MPDD dataset. However, compared with RegAD-L, the proposed RegAD improves a lot, showing the effectiveness of the proposed feature registration aggregated training procedure on multiple categories.

Comparison with Vanilla Anomaly Detection Methods. The state-of-the-art vanilla AD methods use the whole normal dataset for their training and train a separate model for each category, so their performance can be seen as the upper bound for FSAD. We consider methods including GANomaly [1], ARNet [42], MKD [33], CutPaste [21], FYD [45], PaDiM [8], PatchCore [28] and CflowAD [14]. Results in Table 4 show that the proposed RegAD reaches competitive performance even compared with vanilla AD methods that are based on extensive normal data. For example, with only 4 support images, the proposed method (88.2% AUC) outperforms MKD (87.7%) with the same ImageNet pre-trained backbone, and with 32 support images its AUC increases to 94.6%.

5.3 Ablation Studies

Experiments were performed to evaluate the contribution made by individual components of the proposed method. Results of ablation studies for k-shot

Table 5. Ablation studies of k-shot anomaly detection and localization on the MVTec and MPDD datasets. Modules of 'A', 'F', and 'S' mean the augmentations for the support set, the feature registration aggregated training, and the spatial transformer networks (STN), respectively. Results are listed as the macro-average AUC in % over all categories in each dataset of 10 runs. The best-performing method is in bold.

Modules			MVTec						MPDD					
			image			pixel			image			pixel		
A	F	S	k=2	k=4	k=8	k=2	k=4	k=8	k=2	k=4	k=8	k=2	k=4	k=8
			74.7	78.0	80.5	88.6	90.5	92.1	49.6	53.7	55.5	89.5	91.2	92.0
✓			81.5	84.9	87.4	93.3	94.7	95.5	50.8	54.2	61.1	92.4	93.3	93.9
	✓		78.0	80.9	83.1	90.8	92.5	94.0	53.9	55.5	57.2	91.5	92.2	93.0
	✓	✓	79.1	82.9	84.9	90.5	93.3	94.3	57.6	60.9	62.7	91.0	91.8	93.0
✓	✓		83.0	86.4	89.3	**94.7**	**95.9**	96.6	52.8	57.7	64.8	**93.3**	**94.1**	94.4
✓	✓	✓	**85.7**	**88.2**	**91.2**	94.6	95.8	**96.7**	**63.4**	**68.8**	**71.9**	93.2	93.9	**95.1**

Table 6. Ablation studies of different transformation versions of STN modules on MVTec and MPDD for anomaly detection with $k = 2$. T, R means translation, and rotation, respectively. Results are listed as the macro-average AUC in % over all categories in each dataset of 10 runs. The best-performing method is in bold.

Data	no STN	T	R	scale	shear	R +scale	T +scale	T+R	T+R +scale	affine
MVTec	83.0	84.5	85.0	84.9	84.9	**85.7**	84.9	84.2	84.9	84.5
MPDD	52.8	62.3	57.7	59.2	59.0	61.5	61.8	61.0	61.7	**63.4**

anomaly detection and localization on the MVTec and MPDD datasets are shown in Table 5. Modules of 'A', 'F', and 'S' mean the augmentations for support sets, the feature registration aggregated training on multiple categories, and the spatial transformer networks (STN), respectively. Results in Table 5 show that:

(i) Augmentations. The proposed support set augmentations are shown to be essential for both detection and localization. With $k = \{2, 4, 8\}$, the AUC is improved for 6.8%, 6.9%, 6.9% on MVTec and for 1.2%, 0.5%, 0.6% on MPDD, respectively. We further presents the ablation studies of comparing different augmentation methods for support images in the supplementary material.

(ii) Feature Registration Aggregated Training. The feature registration aggregated training on multiple categories is effective both with and without support image augmentations. It shows that the proposed feature registration is beneficial for estimating the normal distribution. As shown in Table 5, with $k = \{2, 4, 8\}$, the proposed anomaly-free feature registration can improve the AUC by 3.3%, 2.9%, 2.6% on MVTec, respectively.

(iii) Spatial Transformer Modules. The proposed STN module is good for improving the ability of the feature registration and thus beneficial for AD. For example, as shown in Table 5, when $k = 8$, the STN module can further improve the performance from 89.3% to 91.2% on MVTec and from 64.8% to 71.9% on MPDD. However, models with STN modules show similar pixel-level localization performance with models without STN modules. The reason comes from the information lost of the inverse transformation operation and its impre-

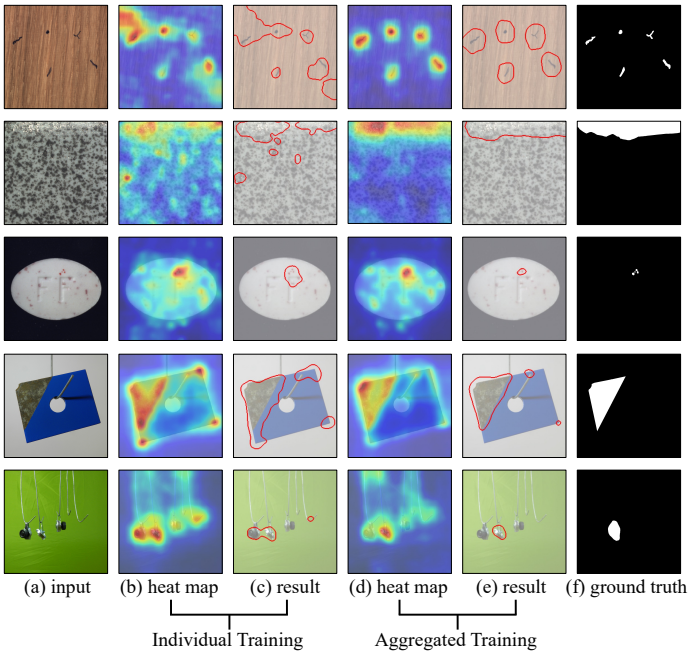

Fig. 3. Qualitative results of anomaly localization for RegAD on the MVTec dataset (top three rows) and the MPDD dataset (bottom two rows) for several cases, including localization results with individual training and aggregated training. Results from (e) show better performance than results from (c), showing the effectiveness of the proposed feature registration aggregated training procedure.

cision. These inverse transformations are designed as post-processing operations to rematch the spatial location of transformed features and the original images.

We further conduct ablation studies on different transformation versions of STN modules on MVTec and MPDD for AD, as shown in Table 6. The best performing STN version is rotation+scale on MVTec, which matches the observation that samples in this dataset are all aligned to the center, and thus, there is no need for translation. While for the MPDD dataset, since the samples are not well be centered, the version of STN with affine transformations shows the best performance. STN is used as a feature transformation module, enabling the model to implicitly transform the images to facilitate feature registration. Images in MPDD are captured under various spatial orientations and positions, thus aligning the features is expected to be helpful. For MVTec, objects are well centralized and have similar orientations, so STN is less helpful to MVTec.

5.4 Visualization Analysis

To qualitatively analyze how the proposed feature registration approach improves the anomaly localization performance, we visualize the results of some

(a) Without Feature Registration (b) With Feature Registration

Fig. 4. Visualization, using t-SNE, of the features learned from the MVTec dataset, using (a) the baseline without the feature registration, and (b) the proposed method with the feature registration. The same t-SNE optimization iterations are used in each case. Results show that features with registration are more compact within each category, and more separated from different categories.

cases from the MVTec and MPDD datasets. It can be seen from the results in Fig. 3 that the localization produced by RegAD using aggregated training (column e) is closer to the ground truth (column f) than that produced by the individual training baseline (column c). This illustrates the effectiveness of the proposed feature registration training procedure on multiple categories.

We also use t-SNE [23] to visualize the features learned on the MVTec dataset, as shown in Fig. 4. Each dot here represents an augmented normal sample from the test set. It can be seen that the proposed feature registration makes the features more compact within each category, and pushes away features of different categories, which is desirable for the benefit of estimating the normal distribution for each category.

6 Conclusion

This paper proposes an FSAD method utilizing registration, a task inherently generalizable across categories, as the proxy task. Given only a few normal samples for each category, we trained a category-agnostic feature registration network with the aggregated data. This model is shown to be directly generalizable to new categories, requiring no re-training or parameter fine-tuning. The anomalies are identified by comparing the registered features of the test image and its corresponding support (normal) images. For both anomaly detection and anomaly localization, the method is shown to be competitive, even compared with vanilla AD methods that are trained with much larger volumes of data. The impressive results suggest a high potential for the proposed method to be applicable in real-world anomaly detection environments.

Acknowledgments. This work is supported by the National Key Research and Development Program of China (No. 2020YFB1406801), 111 plan (No. BP0719010), and STCSM (No. 18DZ2270700), and State Key Laboratory of UHD Video and Audio Production and Presentation.

References

1. Akçay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: Proceedings of the Asian Conference on Computer Vision (ACCV). pp. 622–637. Springer (2018) 4, 8, 11
2. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9592–9600 (2019) 3, 8
3. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: Signature verification using a siamese time delay neural network. Advances in neural information processing systems (NeurIPS) **6** (1993) 6
4. Brown, L.G.: A survey of image registration techniques. ACM computing surveys (CSUR) **24**(4), 325–376 (1992) 3
5. Chen, X., He, K.: Exploring simple siamese representation learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 15750–15758 (2021) 3, 6
6. Chen, Z., Fu, Y., Zhang, Y., Jiang, Y.G., Xue, X., Sigal, L.: Multi-level semantic feature augmentation for one-shot learning. IEEE Transactions on Image Processing **28**(9), 4594–4605 (2019) 4
7. Cheng, M.M., Mitra, N.J., Huang, X., Torr, P.H., Hu, S.M.: Global contrast based salient region detection. IEEE transactions on pattern analysis and machine intelligence **37**(8), 569–582 (2014) 4, 10
8. Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: Proceedings of the IEEE/CVF International Conference on Pattern Recognition (ICPR). pp. 475–489. Springer (2021) 3, 4, 7, 8, 11
9. Eskin, E.: Anomaly detection over noisy data using learned probability distributions. In: International Conference on Machine Learning (ICML) (2000) 4
10. Eykholt, K., Evtimov, I., Fernandes, E., Li, B., Rahmati, A., Xiao, C., Prakash, A., Kohno, T., Song, D.: Robust physical-world attacks on deep learning visual classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1625–1634 (2018) 1
11. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning (ICML). pp. 1126–1135 (2017) 4
12. Golan, I., El-Yaniv, R.: Deep anomaly detection using geometric transformations. In: Advances in neural information processing systems (NeurIPS). vol. 31 (2018) 1, 4
13. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., Hengel, A.v.d.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1705–1714 (2019) 1, 4
14. Gudovskiy, D., Ishizaka, S., Kozuka, K.: Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 98–107 (2022) 4, 8, 11
15. He, J., Hong, R., Liu, X., Xu, M., Wang, M.: Revisiting deep local descriptor for improved few-shot classification. 30th International Joint Conference on Artificial Intelligence (IJCAI) pp. 3420–3426 (2021) 4

16. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 770–778 (2016) 5, 9
17. Huang, C., Xu, Q., Wang, Y., Wang, Y., Zhang, Y.: Self-supervised masking for unsupervised anomaly detection and localization. IEEE Transactions on Multimedia (2022) 4
18. Huang, C., Ye, F., Zhao, P., Zhang, Y., Wang, Y., Tian, Q.: Esad: End-to-end semi-supervised anomaly detection. In: The 32nd British Machine Vision Conference (BMVC) (2022) 1
19. Jaderberg, M., Simonyan, K., Zisserman, A., et al.: Spatial transformer networks. Advances in neural information processing systems (NeurIPS) **28** (2015) 3, 6
20. Jezek, S., Jonak, M., Burget, R., Dvorak, P., Skotak, M.: Deep learning-based defect detection of metal parts: evaluating current methods in complex conditions. In: International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). pp. 66–71. IEEE (2021) 3, 8
21. Li, C.L., Sohn, K., Yoon, J., Pfister, T.: Cutpaste: Self-supervised learning for anomaly detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 9664–9674 (2021) 4, 7, 8, 11
22. Liu, J., Sun, Y., Han, C., Dou, Z., Li, W.: Deep representation learning on long-tailed data: A learnable embedding augmentation perspective. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 2970–2979 (2020) 4
23. Maaten, L.v.d., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008) 14
24. Matsubara, T., Tachibana, R., Uehara, K.: Anomaly machine component detection by deep generative model with unregularized score. In: 2018 International Joint Conference on Neural Networks (IJCNN). pp. 1–8. IEEE (2018) 1
25. Peng, H., Chung, P., Long, F., Qu, L., Jenett, A., Seeds, A.M., Myers, E.W., Simpson, J.H.: Brainaligner: 3d registration atlases of drosophila brains. Nature methods **8**(6), 493–498 (2011) 3
26. Rahmani, M., Atia, G.K.: Coherence pursuit: Fast, simple, and robust principal component analysis. IEEE Transactions on Signal Processing **65**(23), 6260–6275 (2017) 4
27. Ravi, S., Larochelle, H.: Optimization as a model for few-shot learning. In: International Conference on Learning Representations (ICLR) (2017) 4
28. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14318–14328 (2022) 4, 8, 11
29. Rudolph, M., Wandt, B., Rosenhahn, B.: Same same but differnet: Semi-supervised defect detection with normalizing flows. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 1907–1916 (2021) 2, 3, 4, 7, 8, 9, 10
30. Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S.A., Binder, A., Müller, E., Kloft, M.: Deep one-class classification. In: International Conference on Machine Learning (ICML). pp. 4393–4402 (2018) 1, 4
31. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International Journal of Computer Vision **115**(3), 211–252 (2015) 4

32. Sabokrou, M., Khalooei, M., Fathy, M., Adeli, E.: Adversarially learned one-class classifier for novelty detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 3379–3388 (2018) 4
33. Salehi, M., Sadjadi, N., Baselizadeh, S., Rohban, M.H., Rabiee, H.R.: Multiresolution knowledge distillation for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 14902–14912 (2021) 1, 4, 8, 11
34. Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: International Conference on Information Processing in Medical Imaging. pp. 146–157. Springer (2017) 4
35. Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution. Neural computation **13**(7), 1443–1471 (2001) 1, 4
36. Sheynin, S., Benaim, S., Wolf, L.: A hierarchical transformation-discriminating generative model for few shot anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 8495–8504 (2021) 2, 3, 4, 7, 8, 9, 10
37. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. Advances in neural information processing systems (NeurIPS) **30** (2017) 4
38. Sung, F., Yang, Y., Zhang, L., Xiang, T., Torr, P.H., Hospedales, T.M.: Learning to compare: Relation network for few-shot learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 1199–1208 (2018) 4
39. Wu, J.C., Chen, D.J., Fuh, C.S., Liu, T.L.: Learning unsupervised metaformer for anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 4369–4378 (2021) 1, 4, 10
40. Xia, Y., Cao, X., Wen, F., Hua, G., Sun, J.: Learning discriminative reconstructions for unsupervised outlier removal. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 1511–1519 (2015) 4
41. Yang, S., Liu, L., Xu, M.: Free lunch for few-shot learning: Distribution calibration. In: International Conference on Learning Representations (ICLR) (2021) 4
42. Ye, F., Huang, C., Cao, J., Li, M., Zhang, Y., Lu, C.: Attribute restoration framework for anomaly detection. IEEE Transactions on Multimedia **24**, 116–127 (2022) 1, 4, 8, 11
43. Yi, J., Yoon, S.: Patch svdd: Patch-level svdd for anomaly detection and segmentation. In: Proceedings of the Asian Conference on Computer Vision (ACCV) (2020) 1, 4
44. Zhang, J., Xie, Y., Liao, Z., Pang, G., Verjans, J., Li, W., Sun, Z., He, J., Yi Li, C.S.: Viral pneumonia screening on chest x-ray images using confidence-aware anomaly detection. IEEE transactions on medical imaging **40**(3), 879–890 (2021) 1
45. Zheng, Y., Wang, X., Deng, R., Bao, T., Zhao, R., Wu, L.: Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization. arXiv preprint arXiv:2110.04538 (2021) 1, 4, 6, 8, 11
46. Zitová, B., Flusser, J.: Image registration methods: A survey. Image and Vision Computing **21**(11), 977–1000 (2003) 3
47. Zong, B., Song, Q., Min, M.R., Cheng, W., Lumezanu, C., et al.: Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In: International Conference on Learning Representations (ICLR) (2018) 1, 4