FastRecon: Few-shot Industrial Anomaly Detection via Fast Feature Reconstruction

Zheng Fang[1,2†]   Xiaoyang Wang[1,2†]   Haocheng Li[3]   Jiejie Liu[1,3]   Qiugui Hu[3]   Jimin Xiao[1*]

[1]XJTLU   [2]Metavisioncn   [3]Dinnar Automation Technology

{fz.jun26th, haocheng.li789}@gmail.com,
{xiaoyang.wang20, jiejie.liu22}@student.xjtlu.edu.cn, jimin.xiao@xjtlu.edu.cn

Abstract

In industrial anomaly detection, data efficiency and the ability for fast migration across products become the main concerns when developing detection algorithms. Existing methods tend to be data-hungry and work in the one-model-one-category way, which hinders their effectiveness in real-world industrial scenarios. In this paper, we propose a few-shot anomaly detection strategy that works in a low-data regime and can generalize across products at no cost. Given a defective query sample, we propose to utilize a few normal samples as a reference to reconstruct its normal version, where the final anomaly detection can be achieved by sample alignment. Specifically, we introduce a novel regression with distribution regularization to obtain the optimal transformation from support to query features, which guarantees the reconstruction result shares visual similarity with the query sample and meanwhile maintains the property of normal samples. Experimental results show that our method significantly outperforms previous state-of-the-art at both image and pixel-level AUROC performances from 2 to 8-shot scenarios. Besides, with only a limited number of training samples (less than 8 samples), our method reaches competitive performance with vanilla AD methods which are trained with extensive normal samples. The code is available at https://github.com/FzJun26th/FastRecon.

1. Introduction

Anomaly detection (AD) in computer vision aims to identify rare observations that deviate significantly from the majority of the data and do not conform to a well-defined notion of normal behavior [4]. In industrial settings, AD is widely used to detect the anomalies on the surface of the product, which has been attracting a lot of attention re-

---

†Equal contribution.
*Corresponding author.
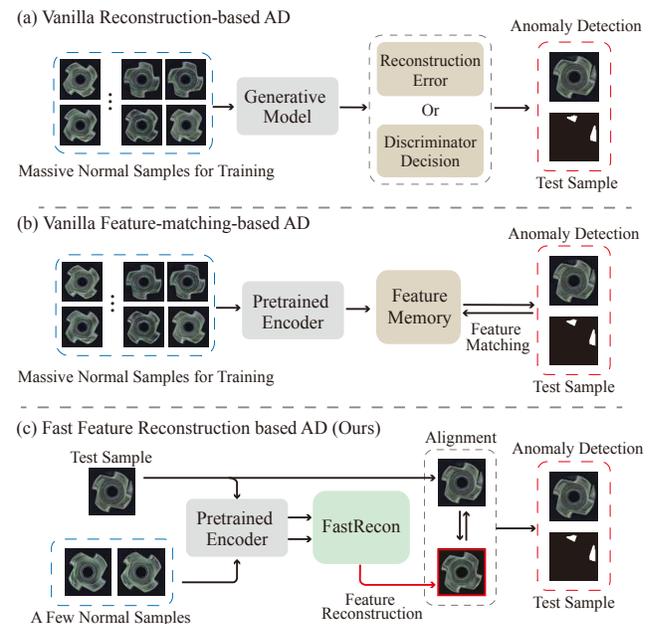
Figure 1: Comparison of our FastRecon with (a) Vanilla Reconstruction-based method and (b) Vanilla Feature-Matching-based method. Compared with (a), our method needs no training session and requires less normal samples for feature reconstruction. Compared with (b), our method applies direct pixel-to-pixel alignment between the test sample and reconstruction result for anomaly detection without an exhaustive feature searching process, thus being more efficient in inference.

cently [31, 16, 9]. The difficulty of industrial AD lies in the diversity of the defects, which can take different forms, from subtle flaws like slight bruises, to obvious ones such as missing components [3]. This makes it difficult to detect and classify anomalies. Primal researches focus on utilizing supervised learning to tackle industrial AD [5, 17]. However, such a learning paradigm often requires an exhaustive set of anomalous samples for each category, which can be

naturally rare to collect, making it inefficient for real-world industrial scenarios.

Recent works pay attention to learning with limited or even no supervision to overcome the problem caused by the lack of anomalous samples. As shown in Fig. 1 (a), reconstruction-based industrial AD [34, 10, 26, 14] utilizes generative models such as autoencoder to learn the distribution of normal samples through the cumbersome training process. As in Fig. 1 (b), recent researches favor feature-matching strategy [28, 7] that conducts feature comparison between the test sample and memory bank of normal features, where the feature searching causes high computational complexity. Both lines of work require massive normal samples either for model training or memory construction, only to achieve detection for a single product category, and fail to quickly generalize across products.

In this paper, we state that well-generalized industrial AD can be achieved in an extremely data-efficient way. We propose a training-free, few-shot AD method named *FastRecon*, to meet the demand for industrial scenarios, which works under the situation that only a limited number of normal samples are available. Given a query sample to detect, we propose to construct its normal version at the feature level through a linear transformation based on features extracted from limited normal samples. This constructed sample shares low-frequency features with the original query image but differs in the defect regions, making the anomaly apparent when directly compared to the original. The comparisons on the methodology of our method and previous mainstream works are shown in Fig. 1. The key lies in finding the optimal transformation. On the one hand, the transformation should bring the constructed sample somewhere close to the query sample so they share visual features on the normal parts. On the other hand, the constructed sample should maintain the property of a normal sample so the difference between the two samples can fully reveal the anomaly regions. With this goal, we proposed a distributional regularization term upon the classical ridge regression algorithm. We first utilize normal features to fit a Gaussian distribution. When finding the transformation from normal features to the query sample, we simultaneously push the reconstructed sample toward the center of the built distribution to enhance its normality. Note that our proposed optimization process admits a closed-form solution just as ridge regression, which allows fast computation with low cost, ensuring the adaptation and migration of our method from old products to novel ones.

Experiments on the benchmark datasets for industrial defection, MVTec [3] and MDPP [15], as well as comparisons with various existing state-of-the-art FSAD methods [19, 13], indicate the effectiveness of our model. Compared with previous state-of-the-art [13], our method achieves improvements of 5.27%, 6.04%, 3.99% on MVTec, and 10.25%, 11.55%, 10.60% on MPDD, at image-level AU-ROC under 2-shot, 4-shot and 8-shot scenarios, respectively. The contributions of this work are summarized below:

1. We propose a few-shot and category-agnostic industrial AD strategy which adopts sample alignment between the original test sample and its reconstructed normal version.

2. We propose a regression algorithm with distribution regularization which has a closed-form solution for the transformation estimation, keeping the reconstructed result close to the query image while maintaining the property of normal samples.

3. Our proposed method achieves state-of-the-art few-shot AD performance under various datasets and settings.

2. Related Work

2.1. Anomaly Detection

Industrial visual anomaly detection (AD) is a rising topic in computer vision. The mainstream AD methods can be categorized as reconstruction-based model and feature-matching-based models. Previous reconstruction-based models classified anomalous samples either though reconstruction loss [2, 33, 30] or via comparison between reconstructed normal samples and test sample [21, 22, 29] with generative models. Another line of works, feature-matching-based models, take full advantage of pre-trained features. For example, SPADE [6] firstly put forward memory banks comprising various feature hierarchies for fine-grained, KNN-based [8] anomaly segmentation and image-level anomaly detection. Similarly, PaDiM [7] and Patch-Core [28] investigated AD performance on MVTec in unsupervised learning settings. However, both two lines of methods show greatly high data dependence which would hinder their industrial application.

2.2. Few-shot Anomaly Detection

Few-shot anomaly detection (FSAD) has developed to meet the demands of fast manufacturing changeover. TDG [23] proposes a hierarchical generative model that captures the multi-scale patch distribution of each support image. DiffNet [19] 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. RegAD [13] explored a new paradigm for FSAD, by learning a common model shared among multiple categories in a meta-learning style. Despite of their decent inference performances under few-shot settings, the training process be-
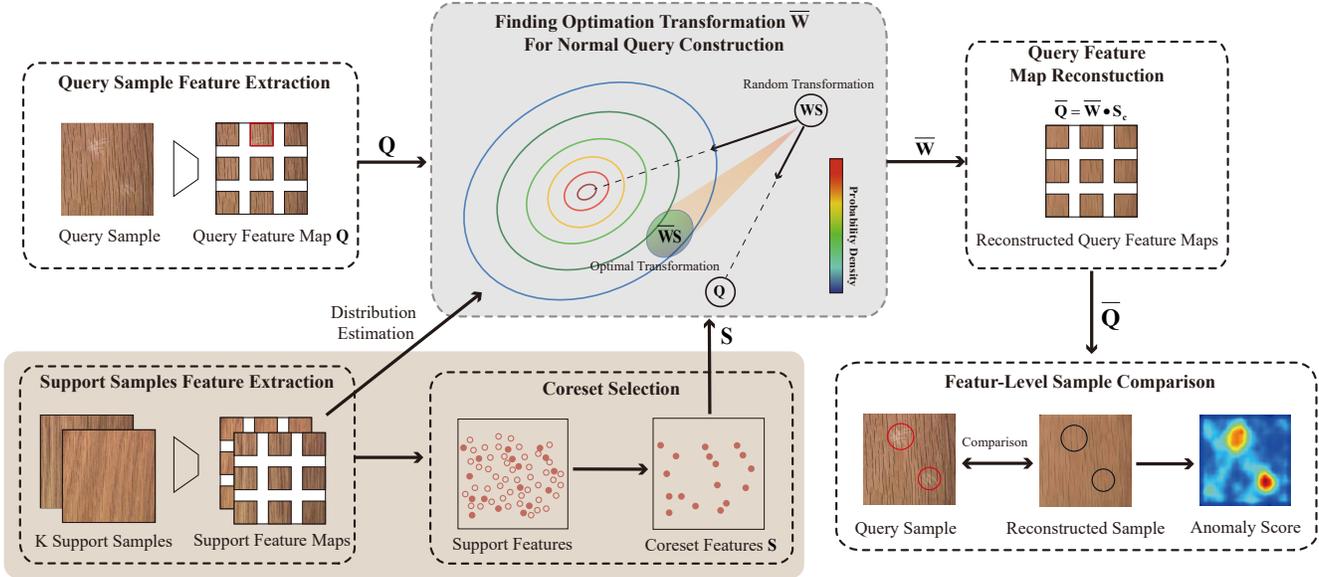
Figure 2: Overview of our method. Feature maps of each query sample and support samples are exacted by a pre-trained encoder. Features from support images are aggregated into a support feature pool. This pool is down-sampled through greedy coreset selection as S to reduce data redundancy and improve inference speed. The coreset S and the original query feature map Q are then input to our proposed regression with distribution regularization as shown in the grey region. An optimal transformation $\bar{W}$ between S and Q is obtained by the regression to make sure the reconstructed sample $\bar{W}S$, denoted as $\bar{Q}$, to share similarity with Q but keeps all the property of normal samples. Finally, we align $\bar{Q}$ and Q for direct comparison to obtaining the anomaly estimation.

comes the main obstacle for fast model migration in real industrial scenarios.

3. Methodology

In this section, we introduce the details of FastRecon. Firstly, we define the one-class classification problem in industrial anomaly detection. FastRecon contains 3 parts: feature extraction and coreset selection, regression with distribution regularization, and anomaly detection with reconstruction. The overview of our method is shown in Fig. 2.

3.1. Problem Definition

We formally define the problem for one-class classification of industrial anomaly detection. Following the few-shot learning setting, in an n-way and k-shot episode, a support set consisting of normal samples from n categories is given, i.e., $S = \bigcup_{i=1}^{n} T_i$, where the subset T_i consists of k normal samples from category c_i. During test time, a query sample, whether normal or anomalous, is presented. The model predicts whether or not the sample is anomalous at the pixel and image level.

3.2. Feature Extraction and Coreset Selection

Existing studies [7, 28] have shown that extracted features from pre-trained CNNs perform well in the anomaly

detection task. We also utilize a pre-trained ResNet-50 to output a patch-level feature map of an image. $\phi_{i,j}$ is used to denote hierarchy-level j of the pre-trained network ϕ for the image $x_i \in X$ (dataset X), where $j \in \{1, 2, 3, 4\}$. We adopt the intermediate features of the hierarchy-level, i.e., $j \in \{2, 3\}$. The resulting feature map has a dimension of $H \times W \times D$, where H, W are the height and width of the largest activation map used to generate the embedding, and D is the total number of channels of both activation maps used to generate embedding.

To facilitate calculations, the feature map is reshaped from 3-dimension $H \times W \times D$ to 2-dimension $R \times D$, where $R = H \times W$. Features in the reshaped map represent embedding vectors, i.e., the local aware patch features. In a k-shot setting, our ultimate support feature pool is obtained by concatenating features of k support images, with a size of $k \times R \times D$. Besides, following [18], we take a coreset sampling operation to get a coreset $S \in R^{c \times d}$, where c is the number of clustering centers computed with K-greedy algorithm. The coverage of S is comparable to the original support feature pools but with a smaller scale so that the computational complexity and the performance of detection could be balanced. For a query image, with the same pre-trained network, it is also extracted to patch-level feature map Q. The process of feature extraction and coreset is

(a) (b) (c) (d)
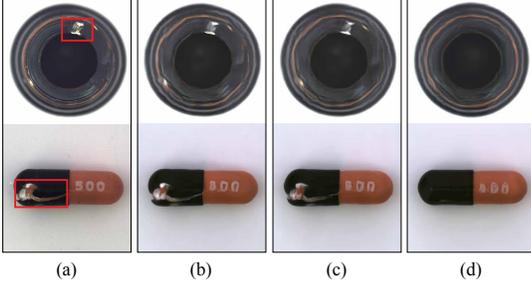
Figure 3: Visual comparison of decoding results on reconstructed features. We train a decoder on features of the whole dataset for image recovery, which is used to probe the transformed features from the encoder. (a) The original query image with defects region in red boxes. (b) Decoding results on original query features. (c) Decoding results on transformed features by ridge regression. The defects regions are reconstructed as on original query features. (d) Decoding results on transformed features by the proposed regression, which shares visual similarity with the original query image but with the property of normal samples.

shown in the brown region in Fig. 2.

3.3. Regression with Distribution Regularization

Why Ridge Regression fails. The goal is to reconstruct query feature map Q as a linear transformation of S-coreset S by finding an optimal transformation weight W such that $WS \approx Q$ and meanwhile WS has most properties of normal samples. Searching for optimal weight W amounts to solving the linear least-squares problem [25]. With original ridge regression, the optimization takes the following form:

$$\bar{W} = \arg\min_{W} \|Q - WS\|^2 + \lambda\|W\|^2, \quad (1)$$

where λ is the weight of the penalty term and $\|\cdot\|$ representing the Frobenius norm. The penalty term in ridge regression, *i.e.*, $\|W\|^2$, ensures tractability when the linear system is over- or under-constrained. However, the transformation obtained from ridge regression can cause the reconstructed sample to deviate significantly from the distribution of normal samples, *i.e.*, after the transformation, the result WS can be too close to Q so that features of defect regions are reconstructed as in Q. As shown in Fig. 3 column (c), the decoding results on reconstructed features share excessively high similarity with the original query image with defects regions being reconstructed.

The distribution regularization. In our work, we propose a distributional regularization term to replace the original penalty term in ridge regression, which effectively enhances the normality of the reconstruction result WS. We first model the distribution of normal feature maps by an isotropic Gaussian density function denoted as $f(X|\mu, \Sigma)$

where μ, Σ is the mean and co-variance estimated by the support set feature maps. We design the regularization term as $|f(WS) - f(\mu)|$ and the proposed regression has the form as:

$$\bar{W} = \arg\min_{W} \|Q - WS\|^2 + \lambda|f(WS) - f(\mu)|, \quad (2)$$

where λ denotes the coefficient that controls the contribution of the regularization. When searching the optimal transformation W, the optimization of the first term $\|Q - WS\|^2$ naturally pushes WS towards original query feature maps Q, while with the proposed regularization $|f(WS) - f(\mu)|$, we force the constructed WS to lie in the high probability density region in the distribution, that is, somewhere close to the distribution center. The visualization of distribution regularization is shown in the grey region of Fig. 2.

The approximation for closed-form solution. The density function inside the regularization term in Eq. (2) prevents us from having a closed-form solution for the regression, which hinders fast calculation of the transformation. Thus, we propose an approximated version of Eq. (2) with the help of Taylor's theorem for multivariable functions on the distribution density function f. Taylor's expansion for $f(x)$ at a is shown as follows:

$$f(x) \approx f(a) + \nabla f(a)(\mathbf{x} - a) + \frac{1}{2}(\mathbf{x} - a)^T H f(a)(\mathbf{x} - a), \quad (3)$$

where $Hf(a)$ denotes Hessian matrix of $f(x)$ at a. And we approximate the density function with the first and second-order term of Taylor's expansion, where the first-order term is 0 at μ. Following Eq. (3), we expand the probability density function $f(WS)$ at μ as:

$$f(WS) \approx f(\mu) + \nabla f(\mu)(WS - \mu)$$
$$+ \frac{1}{2}(WS - \mu)^T Hf(\mu)(WS - \mu). \quad (4)$$

Since our assumed data distribution is isotropic Gaussian, its covariance matrix can be represented as $\Sigma = \sigma^2 I$, where σ^2 is the variance for all dimensions. Then, we have the term $\nabla f(\mu)(WS - \mu)$ as 0, and $Hf(\mu)$ equals the Σ^{-1} which is the inverse of the covariance matrix. Then, by placing $f(\mu)$ to the left, we have

$$f(WS) - f(\mu) \approx \frac{1}{2}(WS - \mu)^T \Sigma^{-1}(WS - \mu)$$
$$= \frac{1}{2\sigma^2}(WS - \mu)^T I(WS - \mu) \quad (5)$$
$$= \frac{1}{2\sigma^2}\|WS - \mu\|^2,$$

where I is an identity matrix. Then we merge the coefficient $\frac{1}{2\sigma^2}$ into λ and the approximated version of the optimization

target in Eq. (2) is:

$$\bar{W} = \arg\min_{W} \|Q - WS\|^2 + \lambda \|WS - \mu\|^2. \quad (6)$$

Finding the optimal transformation. Knowing that the following optimization target of least square error (LSE) of

$$W' = \arg\min_{W} \|Q - WS\|^2, \quad (7)$$

has the closed-form solution as

$$W' = QS^\top \left(S^\top S\right)^{-1}, \quad (8)$$

to utilize the conclusion from Eq. (8), we transform Eq. (6) by variable substitution to fit the form of Eq. (8). Let Q_* and S_* be the column-wise augmented version of Q and S respectively:

$$S_* = \begin{bmatrix} S & \sqrt{\lambda}S \end{bmatrix}, \quad Q_* = \begin{bmatrix} Q & \sqrt{\lambda}\mu \end{bmatrix}, \quad (9)$$

where λ is the coefficient of the regularization term which is always set as a positive number. Hence, equality holds for

$$\|Q_* - WS_*\|^2 = \|Q - WS\|^2 + \lambda \|\mu - WS\|^2, \quad (10)$$

and the form in Eq. (6) can be written as

$$\bar{W} = \arg\min_{W} \|Q_* - WS_*\|^2. \quad (11)$$

Based on the conclusion in Eq. (8), the solution for our optimization process is

$$\begin{aligned} \bar{W} &= Q_* S_*^\top \left(S_*^\top S_*\right)^{-1} \\ &= \left(QS^\top + \lambda \cdot \mu S^\top\right) \left(S^\top S + \lambda \cdot SS^\top\right)^{-1}. \end{aligned} \quad (12)$$

3.4. Anomaly Detection with Reconstruction

After getting the reconstruction query feature maps $\bar{Q}$, we align it with those from the original query sample Q to estimate the pixel-level anomaly segmentation map and image-level anomaly score. For pixel i in the query image, the anomaly score m_i is calculated as:

$$m_i = \|Q_i - \bar{Q}_i\|^2, i \in \mathcal{N}, \quad (13)$$

where $\mathcal{N}$ denotes the collection of all pixel indices in Q. Q_i denotes the relevant feature vector of pixel i. For image-level classification, we use the maximum distance score s^* among all the pixels to represent the image-level score as

$$s^* = \max\{\|Q_i - \bar{Q}_i\|^2 \mid i \in \mathcal{N}\}. \quad (14)$$

Besides, similar to [28, 7], to match the original input resolution, the segmentation map should be up-scaled with bilinear interpolation. Finally, we smoothed the segmentation map with a Gaussian of kernel width of 4.

4. Experiment

4.1. Experimental Setups

Dataset. We evaluated the performance of our model by using the MVTec AD dataset [3] and MPDD [15], which are the real-world benchmarks for AD. MVTec dataset includes 5354 images in 15 categories. 3629 of these images are defect-free and the remaining 1725 have defects. Each category has an average of five different types of defects. The image resolution ranges from 700×700 to 1024×1024. MPDD dataset contains 6 classes of metal productions. Images taken from different distances and spatial directions, as well as the non-uniform backgrounds, make the AD task on this dataset extremely difficult. The MPDD dataset includes 888 normal images in the training set and 176 normal images and 282 abnormal images in the test set with the resolution of 1024×1024. Additionally, both datasets provide ground truth of defective regions on pixel level.

Competing Methods. We compare our method with previous state-of-the-art FSAD and also full-data AD methods. The methods for comparison include GANomaly [1], AR-Net [27], MKD [20], CutPaste [24], FYD [32], PaDiM [7], PatchCore [18], CflowAD [11], TDG [23], DiffNet [19] and RegAD [13].

Evaluation Protocols. We measure the model performance with the area under the Receiver Operating Characteristic curve (AUROC) on classification results. We use image-level AUROC to evaluate the sample detection results and pixel-level AUROC to measure the defects localization performance.

Implementation Details. In our experiment, we used a ResNet-50 [12] model pre-trained on ImageNet for feature extraction. We extract features from the 2 middle-level layers of the model. For both MVTec and MPDD datasets, all the support and query images are reshaped to 368×368 for training and inference. For all the experiments, the coefficient λ for our regularization term is set as 0.3, and the corset selection sampling rate s is set as 0.15. All the experiments are conducted on one NVIDIA GTX 3090 GPU.

4.2. Experiment Results

4.2.1 Comparison with State-of-the-art Methods

Few-shot AD averaged performance. Tab. 1 shows FSAD averaged performance on both MVTec and MPDD datasets. Our method (FastRecon) obtains the best performances under all experimental settings. Specifically, FastRecon improves image-level AUROC by 5.27%, 6.04%, 3.99% on MVTec and 10.25%, 11.55% 10.60% on MPDD over the current state-of-the-art FSAD method RegAD for 2-shot, 4-shot, and 8-shot scenarios respectively. For pixel-level AUROC, FastRecon also outperforms RegAD

Dataset	k	AUROC at Image-level						AUROC at Pixel-level			
		TDG [23]	DiffNet [19]	RegAD [13]	PaDiM [7]	PatchCore [18]	FastRcon (ours)	RegAD [13]	PaDiM [7]	PatchCore [18]	FastRcon (ours)
MVTec	2	71.20	80.60	85.70	78.90	87.81	**90.97**	94.60	90.50	94.75	**95.86**
	4	72.70	81.30	88.20	71.60	89.49	**94.24**	95.80	80.20	94.99	**96.98**
	8	75.20	83.20	91.20	75.30	94.31	**95.19**	96.70	80.50	95.60	**97.27**
MPDD	2	60.30	60.20	63.40	-	59.55	**73.65**	93.20	-	79.15	**97.03**
	4	63.50	63.30	68.30	-	59.78	**79.85**	93.90	-	79.82	**97.60**
	8	68.20	68.50	71.90	-	59.95	**82.50**	95.10	-	80.30	**97.92**

Table 1: FSAD performance comparisons on MVTec and MPDD dataset. The results are averaged over all categories. Both image-level and pixel-level performances are reported in AUROC (%) ↑. k denotes the number of shots in our few-shot settings. The best results for different settings are in bold.

	RegAD [13]		PatchCore [18]		FastRecon	
Category	Image	Pixel	Image	Pixel	Image	Pixel
Bottle	99.40	98.40	**99.60**	**98.60**	99.44	98.50
Cable	76.10	92.70	**97.40**	**97.90**	93.79	96.12
Capsule	72.40	97.60	66.30	97.70	**90.07**	**98.96**
Carpet	97.90	98.90	99.00	99.00	**99.90**	**99.15**
Grid	**91.20**	85.70	63.00	70.60	88.81	**86.32**
Hazelnut	95.80	98.00	92.80	97.00	**99.32**	**98.59**
Leather	**100.00**	99.10	**100.00**	96.90	**100.00**	**99.20**
Metal_Nut	94.60	97.80	94.70	97.00	**99.12**	**98.72**
Pill	80.80	97.40	89.00	96.90	**93.48**	**98.32**
Screw	56.60	95.00	54.10	92.10	**62.46**	**97.10**
Tile	95.50	94.90	**100.00**	96.00	**100.00**	**96.73**
Toothbrush	90.90	98.50	**95.20**	98.80	93.61	**99.04**
Transistor	85.20	93.80	**98.40**	**95.00**	97.29	94.18
Wood	98.60	94.70	97.40	93.10	**99.29**	**94.94**
Zipper	88.50	94.00	95.50	98.30	**96.95**	**98.89**
Average	88.20	95.80	89.49	94.99	**94.24**	**96.98**

Table 2: FSAD categorical performance comparisons in AUROC (%) ↑ on MVTec dataset. Results are from the setting of shot $k = 4$. The best results are in bold. Our method achieves the best performance in most categories.

	RegAD [13]		PatchCore [18]		FastRecon	
Category	Image	Pixel	Image	Pixel	Image	Pixel
Bracket_black	63.80	-	58.90	79.10	**71.81**	**95.54**
Bracket_brown	66.10	-	**70.80**	77.30	63.34	**95.93**
Bracket_white	59.30	-	**70.70**	69.30	69.44	**98.89**
Connector	77.20	-	59.40	86.40	**97.62**	**98.04**
Metal_plate	78.60	-	64.40	86.70	**100.00**	**99.29**
Tubes	67.50	-	34.50	80.10	**76.90**	**97.89**
Average	68.30	93.90	59.78	79.82	**79.85**	**97.60**

Table 3: FSAD categorical performance comparisons in AUROC (%) ↑ on MPDD dataset. Results are from the setting of shot $k = 4$. The best results are in bold. Our method achieves the best performance in most categories.

Method	Data	Backbone	MVTec AD		MPDD	
			Image	Pixel	Image	Pixel
FastRecon	2	WRN50	90.97	95.86	73.65	97.03
FastRecon	4	WRN50	94.24	96.98	79.85	97.60
FastRecon	8	WRN50	95.19	97.27	82.50	97.92
GANomaly [1]	Full	Unet	80.50	-	64.80	-
ARNet [27]	Full	Unet	83.90	-	69.70	-
MKD [20]	Full	Res18	87.70	90.70	-	-
FYD [32]	Full	Res18	97.30	97.40	-	-
PaDiM [7]	Full	WRN50	97.90	97.50	74.80	96.70
PatchCore [18]	Full	WRN50	99.10	98.10	82.10	95.70
CflowAD [11]	Full	WRN50	98.60	98.60	86.10	97.70

Table 4: AD performance comparisons in AUROC (%) ↑ on MVTec and MPDD dataset with full-data methods which utilize the whole dataset in training. Both image-level and pixel-level performances are averaged over all categories. The *Data* column denotes different usage of the dataset.

by 1.26%, 1.18%, 0.57% on MVTev and 3.83%, 3.70%, 2.82% on MPDD respectively.

Few-shot AD categorical performances. Tab. 2 and Tab. 3 show the FSAD categorical performance. As shown in Tab. 2, FastRecon outperforms all the baselines in 10 out of 15 categories at the image level and in 12 out of 15 categories at the pixel level. And as shown in Tab. 3, FastRecon outperforms all the baselines in 4 out of 6 categories at the image level and in all categories at the pixel level.

Comparison with full-data AD. Tab. 4 shows the comparison of our method with AD methods which utilize the whole normal samples. Results show that our FastRecon has competitive performance compared with full-data methods, indicating the high data efficiency of our model.

Speed performance of FastRecon. Speed comparisons on both training and inference are shown in the following Tab. 5. For FastRecon, training stage consists of feature extraction, feature subsampling and calculation on transfor-
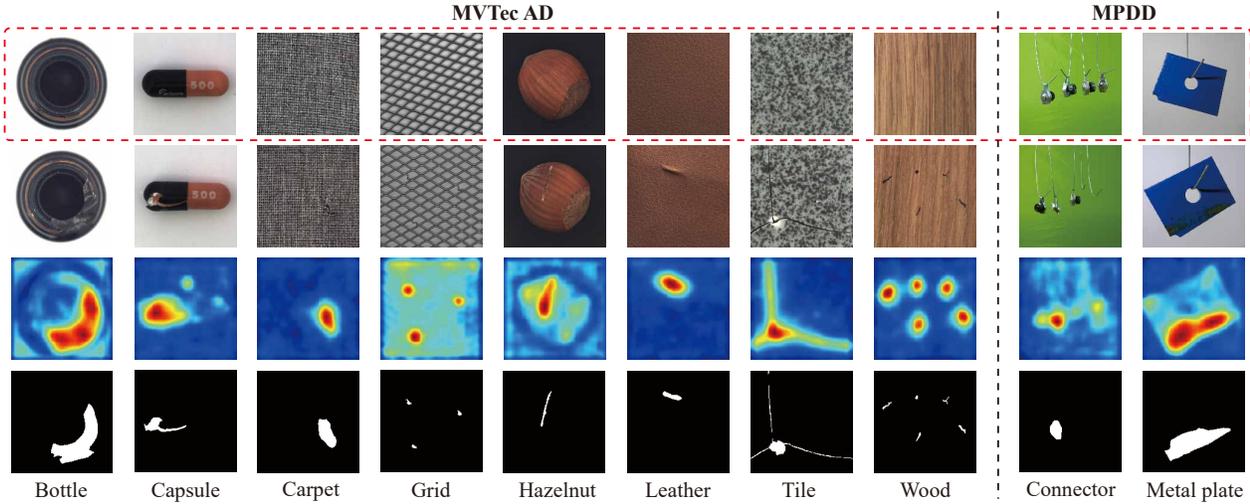
Figure 4: Qualitative results of anomaly localization for both MVTec and MPDD datasets. The first row in the red box presents the support sample for each category while the second row indicates the query samples. The results show that our method can provide accurate localization of defect regions even for more complicated patterns in MPDD.
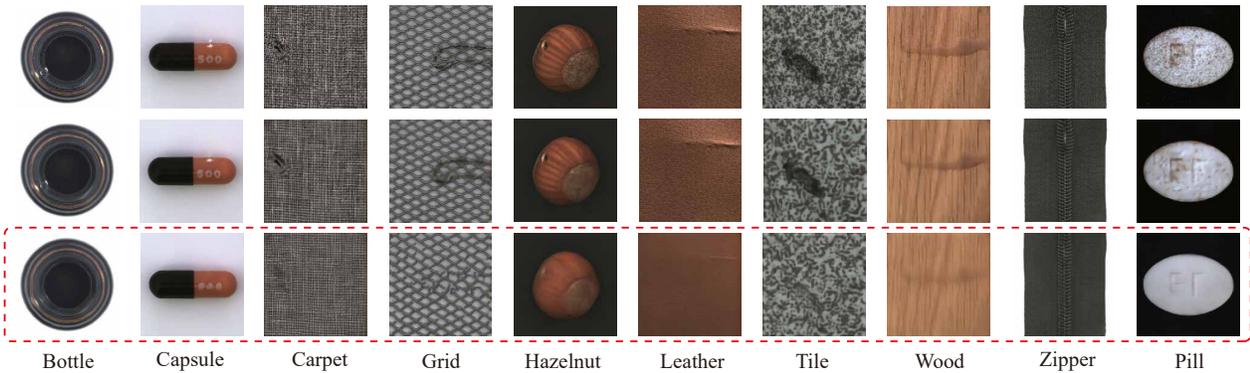
Figure 5: Visualization of decoding results on reconstructed query features. With a pretrained image decoder, we investigate the transformed query features in the MVTec dataset. The first row shows the original query image, and the middle row shows the images decoded from the original query features. The last row represents the image decoded from the reconstructed query feature map. Our method can reveal the normal side of the defective query samples.

	Aggregated Training	Training Time	Inference Time
PatchCore	No	80.37	0.60
RegAD (k=2)	Yes	660.80	0.05
FastRecon (k=8)	No	1.95	**0.04**
FastRecon (k=2)	No	**0.52**	**0.04**

Table 5: Training and inference speed comparisons in seconds. RegAD requires additional training on aggregated features. The inference speed is evaluated as averaged inference time per query sample. FastRecon achieves overall best speed performance with 2 and 8 shots.

mation matrix. FastRecon achieves the best performance.

4.3. Visualization

Localization of anomaly detection Fig. 4 demonstrates visual examples of anomaly detection on MVTec and MPDD. Most defects can be accurately identified.

Decoding results on reconstructed query feature. In order to better reflect the reconstruction ability of FastRecon, a feature decoder is build to visualize the query sample from reconstructed feature maps. The decoder is trained with full MVTec dataset to learn to decode image sample from feature maps. Fig. 5 shows that, images decoded from reconstructed feature maps maintain the normality of the normal samples, which illustrates the effectiveness of our approach.

(a) Sampling Rate s (b) Coefficient λ of Regularization Term   (c) Resolution r
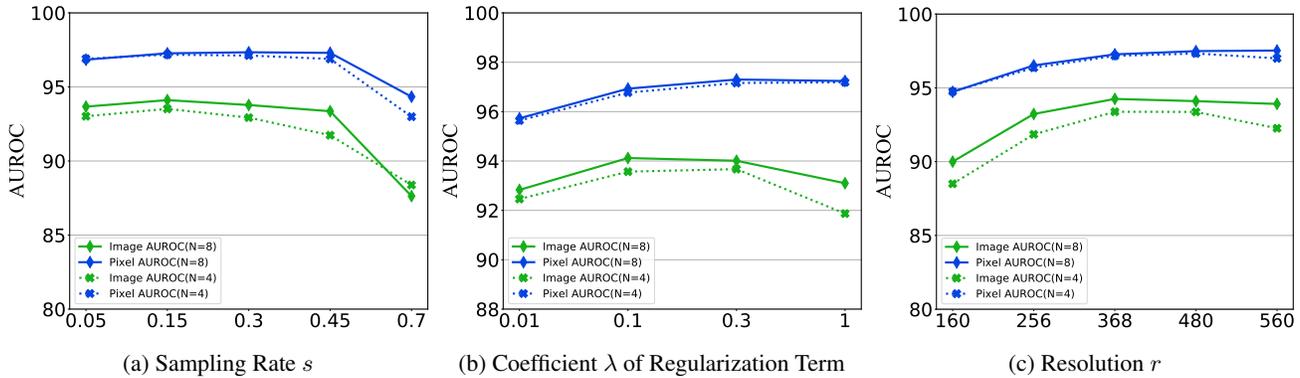
Figure 6: Effect of three hyper-parameters comparisons in AUROC (%) on MVTec dataset. (a) AD performance with different sampling rate s. (b) AD performance with different coefficient λ. (c) AD performance with different resolution r.
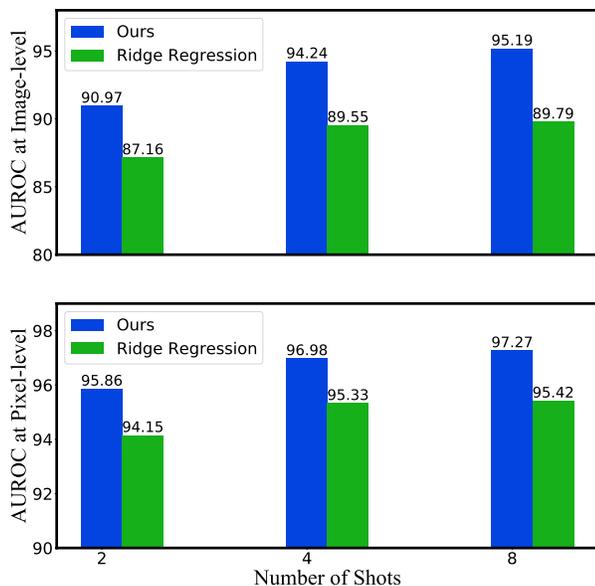
Figure 7: AD performance comparisons in AUROC (%) on MVTec dataset between ridge regression and proposed regularization. Our results, displayed in blue, are better than ridge regression at both image and pixel levels.

4.4. Ablation Studies

In this section, we ablate the core parts of our proposed method to investigate its impact on the final performance. We also discuss the effect of hyper-parameters, which are the coefficient of regularization term λ, the resolution of input data, and the coreset selection sampling ratio s.

Effect of sampling Rate s. Fig. 6 (a) illustrates the relationship between the Sampling Rate s and AUROC performance. The sampling rate s decides the number of feature points in the support feature pool. As shown in Fig. 6 (a), the optimal sampling rate is around 0.45.

Impact of the distribution regularization term and coefficient λ. In Fig. 7, FastRecon obtains better AUROC performance at both image and pixel levels. It reflects the significance of our proposed distribution regularization. Fig. 6 (b) shows the relationship between coefficient λ of regularization term and AUROC performance for 4-shot and 8-shot scenarios. When λ equals 0.3, the AUROC performances are best at both image and pixel levels.

Effect of image resolution r. In our experiment, images, no matter how much the original resolutions are, should be reshaped to new resolution $r \times r$. Fig. 6 (c) reflects the relationship between resolution r and AUROC performance. The resolution will affect the feature points. As shown in Fig. 6 (c), when the resolution equals 368, FastRecon performs well at both image and pixel levels.

5. Conclusion

In this paper, a novel method *FastRecon* is proposed for few-shot anomaly detection based on feature reconstruction. For each query sample, we construct their normal version, referenced by a limited number of normal features. To ensure the reconstructed result shares high visual similarity with the query sample while maintaining the property of the normal sample, a regression equation with proposed distribution regularization is proposed, to obtain the optimal transformation from support to query features. With this fast and accurate reconstruction, the final anomaly detection could be accomplished by a simple alignment. Experimental results show the state-of-the-art performance of our method for 2-shot, 4-shot, and 8-shot scenarios.

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian Conference on Computer Vision (ACCV)*, 2018. 5, 6

[2] Jinwon An and Sungzoon Cho. Variational autoencoder based anomaly detection using reconstruction probability. *Special Lecture on IE*, 2(1), 2015. 2

[3] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. 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)*, 2019. 1, 2, 5

[4] Varun Chandola, Arindam Banerjee, and Vipin Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(15):1–58, 2009. 1

[5] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4):834–848, 2017. 1

[6] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *arXiv preprint arXiv:2005.02357*, 2020. 2

[7] 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 (ICPR)*, 2021. 2, 3, 5, 6

[8] Eleazar Eskin, Andrew Arnold, Michael Prerau, Leonid Portnoy, and Sal Stolfo. A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data. *Applications of Data Mining in Computer Security*, 2002. 2

[9] Kevin Eykholt, Ivan Evtimov, Earlence Fernandes, Bo Li, Amir Rahmati, Chaowei Xiao, Atul Prakash, Tadayoshi Kohno, and Dawn Song. Robust physical-world attacks on deep learning visual classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 1

[10] Dong Gong, Lingqiao Liu, Vuong Le, Budhaditya Saha, Moussa Reda Mansour, Svetha Venkatesh, and Anton van den Hengel. Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 2

[11] Denis Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. 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)*, 2022. 5, 6

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 5

[13] 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 (ECCV)*, 2022. 2, 5, 6

[14] Chaoqin Huang12, Fei Ye, Peisen Zhao13, Ya Zhang, Yanfeng Wang12, and Qi Tian. Esad: End-to-end semi-supervised anomaly detection. In *32nd British Machine Vision Conference (BMVC)*, 2021. 2

[15] Stepan Jezek, Martin Jonak, Radim Burget, Pavel Dvorak, and Milos Skotak. 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)*, 2021. 2, 5

[16] Takashi Matsubara, Ryosuke Tachibana, and Kuniaki Uehara. Anomaly machine component detection by deep generative model with unregularized score. In *International Joint Conference on Neural Networks (IJCNN)*, 2018. 1

[17] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015. 1

[18] 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 (CVPR)*, 2022. 3, 5, 6

[19] Marco Rudolph, Bastian Wandt, and Bodo Rosenhahn. Same same but differnet: Semi-supervised defect detection with normalizing flows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2, 5, 6

[20] Mohammadreza Salehi, Niousha Sadjadi, Soroosh Baselizadeh, Mohammad H Rohban, and Hamid R Rabiee. Multiresolution knowledge distillation for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 5, 6

[21] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Georg Langs, and Ursula Margarethe Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54:30–44, 2019. 2

[22] Thomas Schlegl, Philipp Seeböck, Sebastian M. Waldstein, Ursula Margarethe Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *Information Processing in Medical Imaging (IPMI)*, 2017. 2

[23] Shelly Sheynin, Sagie Benaim, and Lior Wolf. A hierarchical transformation-discriminating generative model for few shot anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2, 5, 6

[24] Kihyuk Sohn, Chun-Liang Li, Jinsung Yoon, and Tomas Jon Pfister. Self-supervised learning for anomaly detection and localization, May 19 2022. US Patent App. 17/454,605. 5

[25] Davis Wertheimer, Luming Tang, and Bharath Hariharan. Few-shot classification with feature map reconstruction net-

works. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 4

[26] Jhih-Ciang Wu, Ding-Jie Chen, Chiou-Shann Fuh, and Tyng-Luh Liu. Learning unsupervised metaformer for anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 2

[27] Fei Ye, Chaoqin Huang, Jinkun Cao, Maosen Li, Ya Zhang, and Cewu Lu. Attribute restoration framework for anomaly detection. *IEEE Transactions on Multimedia*, 24:116–127, 2019. 5, 6

[28] Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2020. 2, 3, 5

[29] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 112:107706, 2020. 2

[30] Houssam Zenati, Manon Romain, Chuan-Sheng Foo, Bruno Lecouat, and Vijay Chandrasekhar. Adversarially learned anomaly detection. In *IEEE International Conference on Data Mining (ICDM)*, 2018. 2

[31] Jianpeng Zhang, Yutong Xie, Guansong Pang, Zhibin Liao, Johan Verjans, Wenxing Li, Zongji Sun, Jian He, Yi Li, Chunhua Shen, et al. Viral pneumonia screening on chest x-rays using confidence-aware anomaly detection. *IEEE Transactions on Medical Imaging*, 40(3):879–890, 2020. 1

[32] Ye Zheng, Xiang Wang, Rui Deng, Tianpeng Bao, Rui Zhao, and Liwei Wu. Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization. In *IEEE International Conference on Multimedia and Expo (ICME)*, 2022. 5, 6

[33] Chong Zhou and Randy C Paffenroth. Anomaly detection with robust deep autoencoders. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2017. 2

[34] Bo Zong, Qi Song, Martin Renqiang Min, Wei Cheng, Cristian Lumezanu, Daeki Cho, and Haifeng Chen. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. In *International Conference on Learning Representations (ICLR)*, 2018. 2