This CVPR paper is the Open Access version, provided by the Computer Vision Foundation.
Except for this watermark, it is identical to the accepted version;
the final published version of the proceedings is available on IEEE Xplore.

Hyperbolic Anomaly Detection

Huimin Li, Zhentao Chen, Yunhao Xu, Junlin Hu*
School of Software, Beihang University, Beijing, China
{lihuimin, hujunlin}@buaa.edu.cn

Abstract

Anomaly detection is a challenging computer vision task in industrial scenario. Advancements in deep learning constantly revolutionize vision-based anomaly detection methods, and considerable progress has been made in both supervised and self-supervised anomaly detection. The commonly-used pipeline is to optimize the model by constraining the feature embeddings using a distance-based loss function. However, these methods work in Euclidean space, and they cannot well exploit the data lied in non-Euclidean space. In this paper, we are the first to explore anomaly detection task in hyperbolic space that is a representative of non-Euclidean space, and propose a hyperbolic anomaly detection (HypAD) method. Specifically, we first extract image features and then map them from Euclidean space to hyperbolic space, where the hyperbolic distance metric is employed to optimize the proposed HypAD. Extensive experiments on the benchmarking datasets including MVTec AD and VisA show that our HypAD approach obtains the state-of-the-art performance, demonstrating the effectiveness of our HypAD and the promise of investigating anomaly detection in hyperbolic space.

1. Introduction

Automatic anomaly detection is one of the key issues in production intelligence [41]. Over the past few years, vision-based automated defect detection has been widely used in semiconductor manufacturing, textile industry, aerospace and other fields, and timely detection of defects is of great significance for quality control in the production process.

Early anomaly detection of industrial image is usually done manually. Since the inherent limitations of the human visual system, it is easy to occur visual fatigue in the face of large-scale data to be detected in a long-term, high-intensity working environment. In this case, as shown in Fig. 1, subtle anomalies such as scratches and cracks are very easy to miss. Hence, automatic anomaly detection based on vision has become the focus of research. Compared with the traditional methods [18, 42, 45], supervised learning-based anomaly detection methods [31] have demonstrated a substantial improvement in both detection accuracy and robustness. Yet, supervised detection methods inevitably require a huge collection of labeled data for the model training to learn effective feature representations, which allows the model an improved generalization capability. However, the actual situation is exactly the opposite, and the anomaly detection task has its intrinsic characteristics and challenges. For instance, the available anomalous data is scarce and it is hardly to collect sufficient defect samples. In addition, the proportion of defective pixels in the image is relatively low, with less available defective features. Furthermore, there are various types and shapes of defects, and we cannot manually describe all defects. In order to solve these problems, self-supervised anomaly detection methods are constantly developed [2, 4, 33]. Recently, teacher-student networks [1, 5, 38] based on knowledge distillation have performed well on publicly available datasets, and image-level detection metrics have been close to saturation. However, pixel-level anomaly detection is still very challenging.

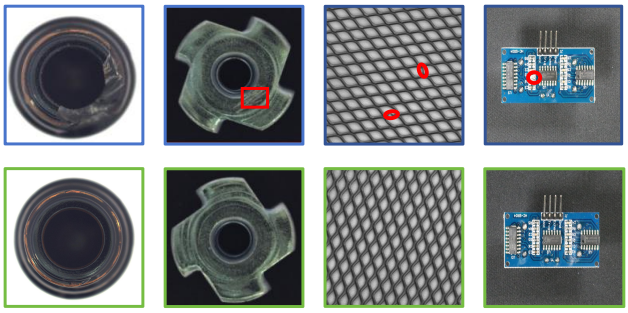


Figure 1. Several examples with/ without anomalies. Images of the first row are anomaly samples, where the red box indicates the region exists subtle anomalies, and the second row showcases their normal samples.

The existing anomaly detection methods are introduced in Euclidean space to measure the image embeddings, as

*Corresponding author.

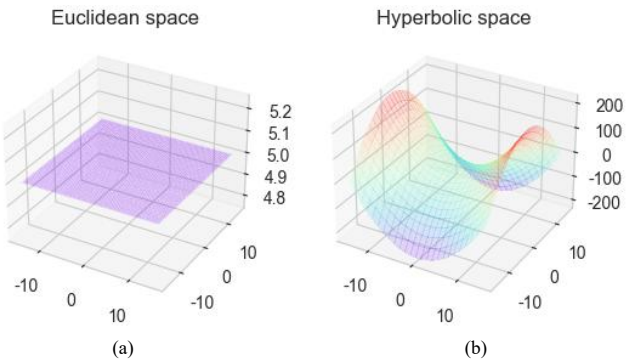


Figure 2. A simple illustration of (a) Euclidean space and (b) hyperbolic space in 3-dimensional space.

shown in Fig. 2(a), and they degrade their performance when the image data lies in non-Euclidean space. Recent work [10, 11, 20] on hyperbolic embeddings has shown that the selection of embedding space directly affects the metric used for comparing feature representations, and that hyperbolic space with negative curvature may be more advantageous in learning feature embeddings, as shown in Fig. 2(b). Based on this finding, we attempt to explore anomaly detection problem in hyperbolic space which is an excellent family member of non-Euclidean space.

In this paper, we propose a hyperbolic space-based anomaly detection (HypAD) method for industrial image. Our HypAD develops the idea of self-supervised anomaly detection methods from Euclidean space into hyperbolic space, and then performs optimization of image embeddings for anomaly localization. To the best of our knowledge, our HypAD is the first method for anomaly detection of industrial image in hyperbolic space. The main contributions of this paper can be summarized as follows:

- We are the first to explore defect detection problem in hyperbolic space, and present a hyperbolic anomaly detection approach, in which the hyperbolic distance metric is utilized to measure the distance between feature representations.
- Extensive experimental results on the MVTec AD and VisA benchmarks show that our HypAD method achieves state-of-the-art performance, which verifies the advantage of our method and the potential of developing anomaly detection via hyperbolic space.

2. Related Work

2.1. Anomaly Detection

Advancements in deep learning constantly revolutionize visual anomaly detection methods in industrial scenario. Previously, anomaly detection tasks usually adopt and improve generalized object detection [32] or segmentation [15, 23, 34] methods by collecting and manually labeling a certain number of defective samples enabling the model to learn a satisfactory defective feature representation. However, these methods struggle in collecting labeled data, making them difficult to be used well in practice.

In recent years, self-supervised anomaly detection methods [2, 4, 5, 22, 27, 33, 49, 51] have developed rapidly. One of the most promising approaches in this category is knowledge distillation-based anomaly detection. Knowledge distillation adopts the Teacher-Student (T-S) paradigm, taking a complex model with a large number of parameters as Teacher and a compact model with a small number of parameters as Student, and aiding the training of the Student model with Teacher. The Teacher is capable of learning and allows it to transfer its knowledge to the Student model to reinforce the capability of the Student. It assumes that a larger regression error exists between the features extracted from student and teacher, thus achieving pixel-level anomaly detection. Initially, Bergmann *et al.* [5] make use of the teacher-student network for anomaly detection, in which the teacher network is frozen during training. Subsequently, various teacher-student networks have been developed for anomaly detection of industrial image [36, 38, 44]. For example, a invertible module is employed as a teacher network in AST [36] to improve the performance of anomaly detection, while MKD [38] adopts a more compact student network, improving inference speed while focusing on distinguishing features. Recently, a lightweight module is designed in EfficientAD [1] to achieve a balance between speed and precision.

However, these methods work in Euclidean space, and they cannot well exploit the image data lied in non-Euclidean space. To overcome this problem, we attempt to extend anomaly detection problem to the hyperbolic space for better exploit the intrinsic geometry of data.

2.2. Hyperbolic Embeddings

Generally, Euclidean space is selected as the embedding space for various visual tasks due to its simple and straightforward concepts, for example, in face recognition and image retrieval tasks [16, 17, 28, 40]. In recent years, several work on natural language processing [29, 30] has reported that learning feature representation in hyperbolic space achieves better performance than that in Euclidean space. Subsequently, some studies have extended the applications of hyperbolic space to visual tasks [10, 12, 20, 24]. Hyperbolic space is a non-Euclidean geometric space that differs from the familiar Euclidean space. In hyperbolic space, the volume grows exponentially relative to the radius, which makes hyperbolic space inherently superior in embedding hierarchical data. Khrulkov *et al.* [20] introduce hyperbolic image embedding method, in which the output features of the model are projected into the hyperbolic space for calculation. Ermolov *et al.* [10] combine

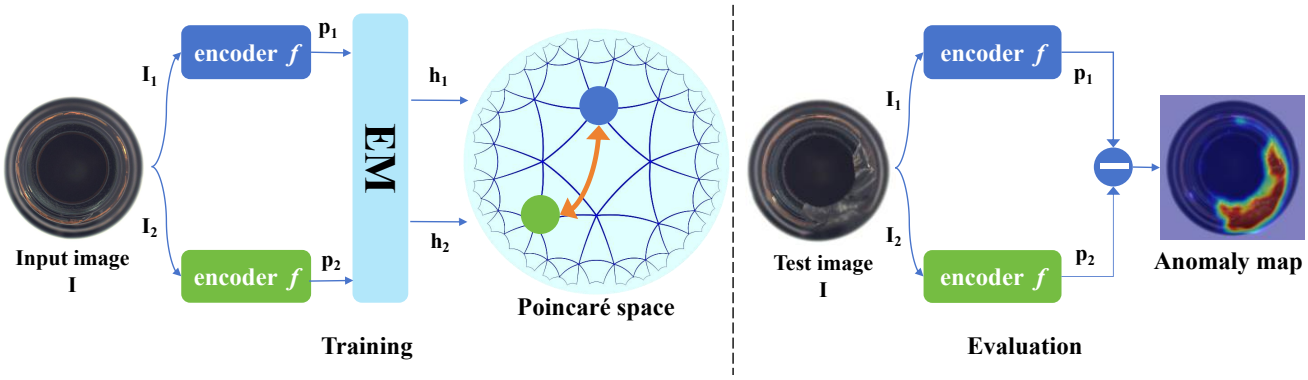


Figure 3. The architecture of our HypAD approach. During training, the two augmented views of an image are processed by encoder *f*, and then the features are mapped into hyperbolic space for further optimization. For evaluation, anomaly map is determined by the difference between image embeddings. EM is an exponential mapping module that maps feature representations from Euclidean space to Poincaré space that is a representative of hyperbolic space.

hyperbolic space with ViT architecture [9] and prove that the performance of hyperbolic corresponding terms is superior to Euclidean settings. Yan *et al.* [46] propose a hyperbolic unsupervised deep metric learning method for image retrieval task, which can effectively capture the hierarchical structure of data.

However, in the field of anomaly detection of industrial image, there is no related research work based on hyperbolic space. To fill this gap, we attempt to explore anomaly detection problem in hyperbolic space and propose a hyperbolic anomaly detection method, which breaks the limitation of the flatness of the Euclidean space on the expressiveness of the data and achieves better results.

3. Proposed Method

This section presents our HypAD method, a new anomaly detection approach in hyperbolic space. Fig. 3 shows the architecture of our HypAD. Our method consists of two processes, training and evaluation. During training, the model is optimized by minimizing the distance between different augmented views of an image, but unlike existing anomaly detection methods, we impose constraints on image embeddings and calculate distance in hyperbolic space. For the evaluation process, anomaly map are obtained by the difference in image embeddings generated by the two branches.

3.1. Preliminary and Motivation

Unlike Euclidean space with zero curvature, the n-dimensional hyperbolic space $\mathbb{H}^n$ is a Riemannian manifold of constant negative curvature and several well-established isometric models of hyperbolic space are available. Following previous work [24], we select Poincaré ball model $(\mathbb{D}_c^n, g^{\mathbb{D}})$ with the curvature parameter c as a representative model of hyperbolic space due to its excellent characteristics. It is noted that the actual curvature value is $-c^2$ and $c > 0$. This Poincaré model is defined by the n-dimensional manifold $\mathbb{D}^n = \{\mathbf{x} \in \mathbb{R}^n : c||\mathbf{x}||^2 < 1\}$ equipped with the Riemannian metric $g^{\mathbb{D}} = \lambda_c^2 g^{\mathbb{E}}$, in which conformal factor $\lambda_c = \frac{2}{1-c||\mathbf{x}||^2}$ and Euclidean metric tensor $g^{\mathbb{E}} = \mathbf{I}_n$, respectively.

Hyperbolic space is not a vector space. That is, we cannot directly use addition and other operations as in Euclidean space. Following the previous practice [10, 11, 20], we utilize the addition of Gyrovector space to define the addition between $\mathbf{x} \in \mathbb{D}_c^n$ and $\mathbf{y} \in \mathbb{D}_c^n$ as:

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{(1 + 2c\langle \mathbf{x}, \mathbf{y} \rangle + c\|\mathbf{y}\|^2)\mathbf{x} + (1 - c\|\mathbf{x}\|^2)\mathbf{y}}{(1 + 2c\langle \mathbf{x}, \mathbf{y} \rangle + c^2\|\mathbf{x}\|^2\|\mathbf{y}\|^2)}. \quad (1)$$

Then, the hyperbolic distance of $\mathbf{x} \in \mathbb{D}_c^n$ and $\mathbf{y} \in \mathbb{D}_c^n$ is measured by:

$$d_{\mathbb{H}}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{c}} \mathsf{arctanh}(\sqrt{c}\| - \mathbf{x} \oplus_c \mathbf{y}\|). \quad (2)$$

Obviously, the distance formula Eq. (2) degenerates into Euclidean distance $d_{\mathbb{E}}$ when c reaches to zero:

$$\lim_{c \to 0} d_{\mathbb{H}}(\mathbf{x}, \mathbf{y}) = 2\|\mathbf{x} - \mathbf{y}\| = d_{\mathbb{E}}(\mathbf{x}, \mathbf{y}). \quad (3)$$

In this paper, we figure that there are hierarchical relationships like tree among most images in the anomaly detection task. As shown in Fig. 4, the whole image is regarded as the root of the tree, while the local area is treated as the node of the tree, and in turn, the smaller local area containing more details can be considered as the deeper node of the tree, which stimulates the inherent advantage of hyperbolic space in dealing with this type of image data. Based on this assumption, we attempt to perform anomaly detection task in hyperbolic space, to effectively exploit hierarchical characteristic of data.

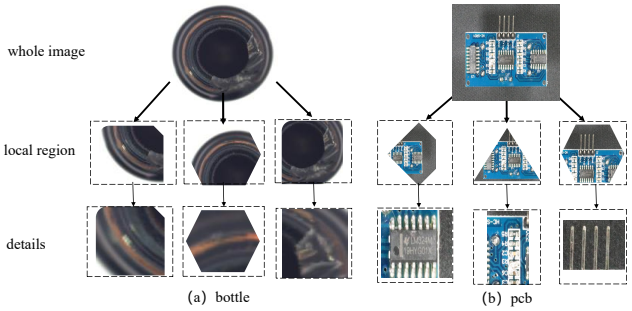


Figure 4. Visual examples of hierarchical relationships in images for anomaly detection task. (a) bottle and (b) pcb are two examples from MVTec AD and VisA datasets, respectively. Here, we want to learn image embeddings that obey the hierarchical constraints. The whole image can be seen as the root of the tree, the local region is treated as the tree node, and in turn, the smaller local area enjoying the more detailed information is considered as the deeper tree node. Hyperbolic space is very suitable for embedding data with such hierarchical structure.

3.2. HypAD

As in Fig. 3, our HypAD takes as input two augmented views $\mathbf{I}_1$ and $\mathbf{I}_2$ from an image $\mathbf{I}$. The two views are first processed by encoder f consisting of a backbone and a simple projection module, and are then mapped into hyperbolic space (*i.e.*, Poincaré space) via exponential mapping (EM) for model optimization.

Backbone. A number of well-established backbones [14, 39, 43] are available for computer vision tasks and recent work [1] has shown that a specific size of receptive field may be helpful in anomaly detection of industrial image. PDN [1] consisting of only a few convolutional and pooling layers is prioritized as the baseline setting due to its excellent performance. With tailored design, each output neuron of the PDN has a fixed pixel receptive field, which may be important for anomaly detection. Similar to ResNet [14], PDN is fully convolutional and can be applied to variable-size images, generating all feature vectors in a single forward pass. PDN-S [1] is a lightweight version of PDN with only four convolutional layers. We use a distilled version of the PDN-S pre-trained on ImageNet as backbone and minimize the mean square error between its output PDN and the features extracted from the pre-trained WideResNet-101 [48] network. This makes it possible to fairly compare with previous methods based on PDN encoder. In particular, features extracted by the backbone are reduced in dimension by a simple projection module (*i.e.* a pooling layer) before being mapped into hyperbolic space.

EM Module. To optimize our HypAD in hyperbolic space, we need to find a mapping function from Euclidean space to hyperbolic space. Exponential mapping (EM) is usually adopted to achieve this aim, which is denoted as Eq. (4) for the fixed base point $\mathbf{h}_0 \in \mathbb{D}_c^n$,

$$\begin{aligned}\mathbf{h} &= \exp_{\mathbf{h}_0}^c(\mathbf{p}) \\ &= \mathbf{h}_0 \oplus_c \left(\tanh\left(\sqrt{c}\frac{\lambda_{\mathbf{h}_0}^c\|\mathbf{p}\|}{2}\right)\frac{\mathbf{p}}{\sqrt{c}\|\mathbf{p}\|}\right),\end{aligned} \tag{4}$$

where $\mathbf{h}_0$ can be simply set to $\mathbf{0}$ as in [10], $\lambda_{\mathbf{0}}^c = 2$, $\mathbf{h} \in \mathbb{D}_c^n$ and $\mathbf{p} \in \mathbb{R}^n$. In addition, logarithmic mapping is termed the inverse mapping of EM from hyperbolic space to Euclidean. In this paper, we only calculate the anomaly loss in the hyperbolic distance, and do not involve the operation of projecting the feature embeddings from the hyperbolic space to the Euclidean space. More details about logarithmic mapping can refer to [20].

Optimization. To optimize our HypAD model, we need to design a suitable loss function in hyperbolic space. Generally, self-supervised contrastive learning in anomaly detection task consists of two forms: one-class anomaly detection using only normal images, and two-class anomaly detection using both normal and defective images. Accordingly, a tailored loss function is performed to optimize the model. For example, a hard feature loss is proposed in [1], whose core idea is to compare image embeddings extracted by the T-S module, and then detect anomalies based on feature similarity, with back propagation taking into account only the most dissimilar local regions within the threshold range. SPD [51] is designed with different augmentation strategies, which narrows the distance between positive samples by choosing an appropriate loss function and further enlarges the distance of negative samples, forcing the model to focus on local anomalous pixels in an image.

Denoting the two output vectors as $\mathbf{p}_1 \triangleq f(\mathbf{I}_1)$ and $\mathbf{p}_2 \triangleq f(\mathbf{I}_2)$, where f is the encoder, the previous methods compute their squared Euclidean distance as:

$$d_{\mathbb{E}}^2(\mathbf{p}_1, \mathbf{p}_2) = ||\mathbf{p}_1 - \mathbf{p}_2||^2. \tag{5}$$

In our work, since we need to measure the distance in hyperbolic space, we cannot directly use the conventional loss function based on Euclidean space. Here, we calculate their hyperbolic distance using Eq. (2), obtaining

$$d_{\mathbb{H}}(\mathbf{h_1}, \mathbf{h_2}) = \frac{2}{\sqrt{c}}\text{arctanh}(\sqrt{c}\|-\mathbf{h_1} \oplus_c \mathbf{h_2}\|), \tag{6}$$

where c is the curvature parameter, $\mathbf{h_1}$ and $\mathbf{h_2}$ are image embeddings in hyperbolic space.

The loss function of our HypAD given by Eq. (6) is straightforward. We want to bring the representations of similar objects in the embedding space closer together. As the commonly-used ways in self-supervised learning-based methods, the encoder is required to yield consistent outputs for different parts of an image obtained with augmentations. This training scheme is in line with the two-class anomaly detection task; in this case, the encoder is explicitly trained

to generate similar output for semantically similar input. It is noted that we use one-class anomaly detection setup, that is, we train only on normal images, thus in hyperbolic space we do not consider the distance between different classes, and we only minimize the distance between feature embeddings of different augmented views of the same image.

4. Experiments

In this section, we empirically evaluate the effectiveness of our HypAD method on the benchmark datasets.

4.1. Datasets and Evaluation Metrics

4.1.1 Datasets

MVTec AD [3] is launched in 2019 and consists of 5354 industrial defect images for 15 categories, of which 3629 images are used for training and validation and 1725 images for testing. It contains five different texture objects and ten different types of objects, which is a commonly used dataset in the field of anomaly detection. **VisA** [51] dataset is released in 2021 and contains 12 subsets totalling 10,821 images with 9,621 normal and 1,200 abnormal samples. There are four subsets that include different types of the printed circuit boards (PCB) with relatively complex in structure, four subsets of objects that are single-class instances, and four subsets of objects that contain multiple instances. Furthermore, MVTec AD and VisA contain various surface defects (*e.g.* crack, scratch, dent, and discolouration) and structural defects (*e.g.* missing parts and misaligned).

4.1.2 Evaluation Metrics

The evaluation metrics of anomaly detection methods consider two main aspects: the detection and localization of anomalies. In [3, 21, 51], the anomaly detection performance is measured with Area Under the Receiver Operator Characteristic curve (AUROC) and Area Under the Precision-Recall curve (AUPR) based on its predicted image-level anomaly scores, while the anomaly localization performance is additionally measured with Area Under the Per-Region Overlap (AUPRO). In particular, the evaluation metrics of the anomaly segmentation approaches are discussed in detail in [3].

4.2. Implementation Details

We use the Adam optimizer for training. The batch size is 1 by default. The learning rate is 0.0001 and it has a step decay schedule. For consistency with previous literature, we also add an additional Autoencoder to get the global reconstruction features. Following [1], we employ the following strategies for data enhancement: Resize, Normalize and ColorJitter. We first resize the input image to 256×256 pixels and then use Normalize(mean = [0.485, 0.456, 0.406], std = [0.229, 0.224, 0.225]) operation. ColorJitter is used to randomly enhance the brightness, contrast, and saturation of an image with a probability of 0.2, *i.e.* RandomChoice([ColorJitter(brightness=0.2), ColorJitter(contrast=0.2), ColorJitter(saturation=0.2)]). Subsequently, the augmented views are fed into the encoder f. ColorJitter is disabled during evaluation. Noting that the backbone has poor generalization ability when using PDN directly. A good practice is to switch to a deeper network or more training data, but this increases the training and reasoning overheads. Here, we follow the settings in [1, 5] and employ distillation to enhance its generalization performance. Hence, an additional training signal is provided by teacher-student distillation with a CNN-based teacher [48], pretrained on ImageNet [37] without labels, which contains approximately 1 million training images of 1000 categories. The encoder f outputs a representation with dimension of 384, followed by a module for feature dimensionality reduction and is then mapped to a Poincaré space with the curvature parameter $c = 0.01$ by default. All our experiments are performed on a single NVIDIA GeForce RTX 3090 GPU.

Several studies [10, 20] have shown that optimizing neural networks in hyperbolic space may suffer from gradient vanishing. A common scheme to guarantee numerical stability is to conduct paradigm clipping after applying the exponential mapping. To alleviate this problem, we utilize the feature clipping technique to ensure numerical stability during training by following the practices in [10], where clipping radius r is set 2.3 by default. Above setup makes it possible to fairly compare with previous work, for this reason, we adopt this configuration in our experiments.

4.3. Results and Analysis

Tab. 1 highlights the experimental results for the pixel-level mean AUPRO performances of current state-of-the-art methods on MVTec AD and VisA datasets. The results for each subset are shown in more detail in Tab. 2 and Tab. 3, respectively. As shown in Tab. 1, our HypAD achieves the

Table 1. Comparison of mean AUPRO scores (%) of current state-of-the-art methods on MVTec AD and VisA datasets.

Method	MVTec AD	VisA
RDAD [7]	93.9	-
S-T [5]	92.4	93.0
FastFlow [47]	92.5	86.8
PatchCore [35]	92.7	79.7
EfficientAD [1]	93.1	93.1
DiffusionAD [50]	-	93.2
FAIR [25]	94.0	91.4
SimpleNet [1]	89.6	68.9
HypAD (Ours)	**97.6 (+3.6)**	**95.4 (+2.2)**

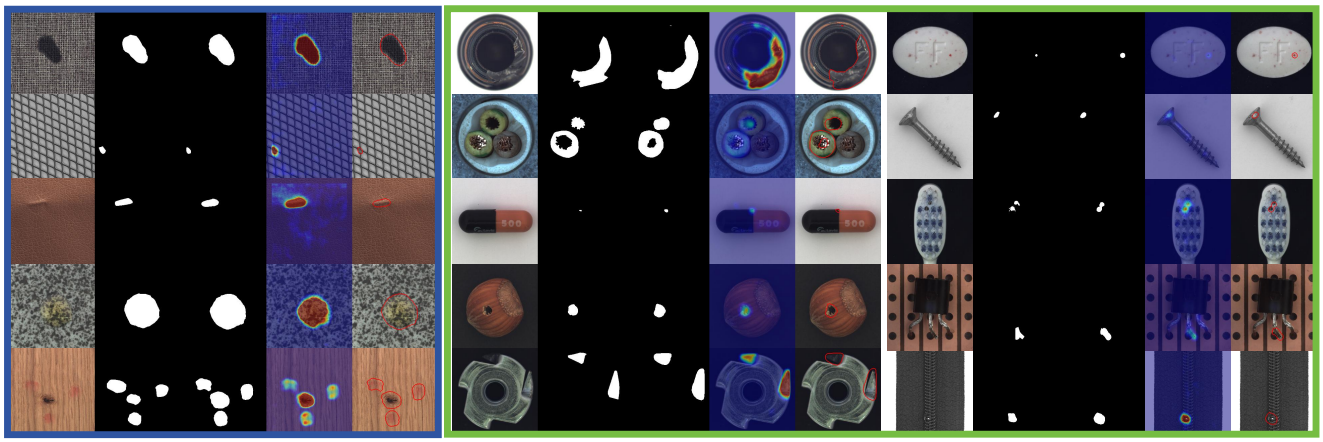
Table 2. Pixel-level AUPRO scores (%) of current state-of-the-art methods on MVTec AD dataset. In this table, "texture Mean" and "object Mean" denote the mean of texture and object categories respectively.

	Category	HypAD (Ours)	S-T [5] CVPR 2020	FCDD [26] ICLR 2021	RDAD [7] CVPR 2022	PyramidFlow [21] CVPR 2023	EfficientAD [1] arXiv 2023
texture	carpet	92.7	87.9	99.0	97.0	97.2	91.7
	grid	99.7	95.2	95.0	97.6	-	88.7
	leather	99.9	94.5	99.0	99.1	99.2	98.2
	tile	99.8	94.6	98.0	90.6	97.2	85.8
	wood	95.3	91.1	94.0	90.9	97.9	89.6
	texture Mean	97.5	92.7	97.0	95.0	97.9	90.8
object	bottle	100	93.1	96.0	96.6	95.5	95.2
	cable	93.3	81.8	93.0	91.0	90.3	89.9
	capsule	96.9	96.8	95.0	95.8	98.3	97.6
	hazelnut	99.7	96.5	97.0	95.5	98.1	95.1
	metal nut	98.0	94.2	98.0	92.3	-	94.1
	pill	98.4	96.1	97.0	96.4	96.1	96.4
	screw	95.6	94.2	93.0	98.2	-	96.1
	toothbrush	99.9	93.3	95.0	94.5	97.9	94.3
	transistor	100	66.6	90.0	78.0	94.7	91.0
	zipper	94.7	95.1	98.0	95.4	95.4	93.2
	object Mean	97.6	90.8	95.2	93.4	95.8	94.3
	Mean	**97.6**	91.4	95.8	93.9	96.5	93.1

Table 3. Pixel-level AUPRO scores (%) of current state-of-the-art methods on VisA dataset.

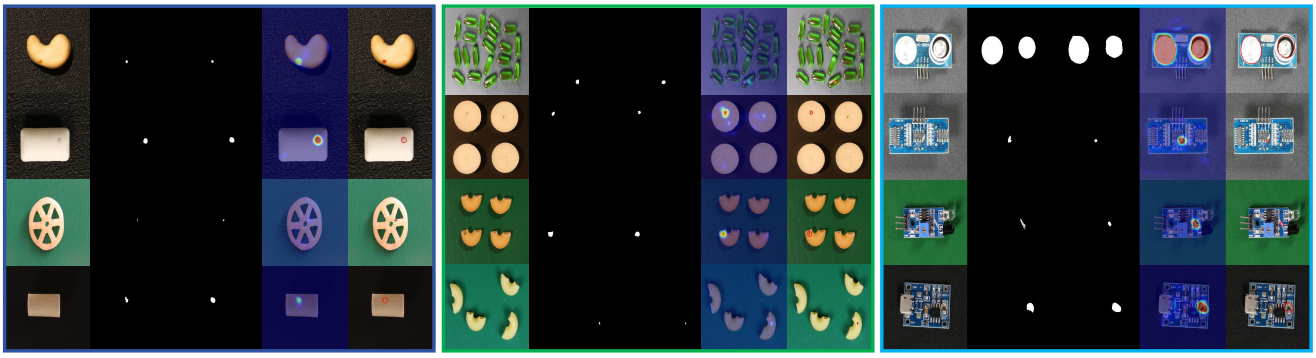
	Category	HypAD (Ours)	S-T [5] CVPR 2020	PatchCore [35] CVPR 2022	RDAD [7] CVPR 2022	DiffusionAD [50] arXiv 2023	EfficientAD [1] arXiv 2023
complex structure	pcb1	97.3	-	94.3	43.2	96.9	96.9
	pcb2	98.7	-	89.2	46.4	92.8	93.8
	pcb3	97.1	-	90.9	80.3	94.4	94.1
	pcb4	97.6	-	90.1	72.2	95.5	91.3
multiple instances	candle	94.9	-	94.0	92.2	94.7	95.0
	capsules	83.1	-	85.5	56.9	97.6	93.3
	macaroni1	98.8	-	95.4	71.3	96.8	99.0
	macaroni2	90.2	-	94.4	68.0	98.0	97.9
single instance	cashew	95.4	-	94.5	79.0	88.0	93.7
	chewing gum	99.6	-	84.6	92.5	87.0	81.5
	fryum	93.8	-	85.3	81.0	96.8	85.9
	pipe fryum	99.6	-	95.7	68.3	80.2	93.8
	Mean	**95.4**	93.0	91.2	70.9	93.2	93.1

best AUPRO scores and outperforms the existing methods by increasing the gains of 3.6% and 2.2% on MVTec AD and VisA respectively, and the similar improvements for most subsets present in Tab. 2 and Tab. 3, which empirically shows that the feature representations in hyperbolic space play a significant role in the problem of anomaly detection of industrial image. Furthermore, these experimental results demonstrate that our HypAD method breaks the limitation of the flatness of the Euclidean space on the expressiveness of the image data. Moreover, the experimental results effectively support our conjecture that industrial images including its augmented views have a certain hierarchical relationships in their internal geometric structures, and such hierarchical characteristic contributes to the anomaly detection task. Our HypAD method captures this relationship, thus achieving better results. In addition, we visualize the original images, ground truths, predict masks, anomaly maps, and final results of each category in MVTec AD and



(a) textures (b) objects

Figure 5. Visualization of each category in MVTec AD. (a) shows five types of texture objects. (b) contains ten categories of objects, three of which (*i.e*. mesh, screw, and zipper) are available only as single-channel images. The visual examples of each class consist of five parts, from left to right are the original image, ground truth, predict mask, anomaly map, and final results. The threshold used by the predict mask is adaptively calculated according to the dataset, without any additional processing.



(a) single instance (b) multiple instance (c) complex structure

Figure 6. Visualization of each category in VisA. (a) includes four types of single instance. (b) contains four types of multiple instances. (c) contains four types of complex structures. The visual examples of each class consist of five parts, from left to right are the original image, ground truth, predict mask, anomaly map, and final results. The threshold used by the predict mask is adaptively calculated according to the dataset, without any additional processing.

VisA datasets, as shown in Fig. 5 and Fig. 6, further validating the advantages of the proposed HypAD.

4.3.1 Effect of Curvature Parameter c

Tab. 4 presents the AUPRO score according to the curvature value c of our HypAD in hyperbolic space on VisA dataset. From the results, it can be seen that larger c value gives rise to degradation. Theoretically, if c tends to zero, then the corresponding radius tends to infinity, making the hyperbolic space as flat as the Euclidean space. In contrast, increasing the value of c means steeper configurations. This indicates that the flatness of Euclidean space may not provide sufficient expressive capability, thus limiting the utilization of the data. Moreover, we find that our HypAD achieves the best performance when $c = 0.01$, so we select this value in our experiments. When $c = 0$, hyperbolic space degenerates into Euclidean space, that is, our HypAD works in Euclidean space without the hyperbolic mapping, and EfficientAD [1] is a special case of our HypAD method.

4.3.2 Evaluation on Anomaly Detection

We then conduct a comparison of image-level defect detection performance of different methods. Following the common settings, the image-level anomaly score is determined

Table 4. Effect of the curvature parameter c in HypAD on VisA.

c	0	0.01	0.05	0.1	0.3	0.5
AUPRO (%)	93.4	**95.4**	95.2	95.2	95.3	95.3

Table 5. Comparison (%) of image-level anomaly detection of various methods on MVTec and VisA datasets.

Method	MVTec AD		VisA	
	AUROC	AUPR	AUROC	AUPR
DRA [8]	95.9	-	-	-
MKD [38]	87.7	-	-	-
SPD [51]	94.6	97.5	87.8	88.6
RDAD [7]	98.5	-	-	-
PatchCore [35]	98.7	98.9	94.3	95.2
SoftPatch [19]	98.6	-	-	-
FAIR [25]	98.6	-	96.7	-
EfficientAD [1]	99.0	98.7	97.6	97.5
DiffusionAD [50]	-	-	97.8	-
SimpleNet [1]	98.2	98.5	87.9	90.1
HypAD (Ours)	**99.2**	**99.5**	**98.3**	**98.5**

Table 6. Image-level AUPRO on MNIST-C in ADBench [13].

Method	DevNet [13]	EfficientAD	HypAD
AUROC (%)	88.04	90.96	**92.51**

by the pixel with maximum value in the anomaly map. As shown in Tab. 5, the results on MVTec AD and VisA show only a slight improvement in image-level performance. The reason is that image-level anomaly detection is relatively simple and its performance has nearly saturated.

To further demonstrate the effectiveness of our HypAD method, we also conduct evaluation on MNIST-C, one of five CV datasets in ADBench [13]. Following the settings on MNIST-C [13], we select the category with labeled anomaly ratio 1% for comparison, as shown in Tab. 6. Our HypAD method obtains the best performance. The AUROC of DevNet [13] is took from Table D6 in supplementary material of ADBench [13], in which DevNet [13] achieves the best result among 16 methods. In future work, we will evaluate our method on other ADBench CV datasets.

4.3.3 Evaluation on Anomaly Localization

Image-level anomaly detection is only a rough judgment of whether there is a defect in the image, and can not really reflect the exact location information of the defect. In this subsection, we report pixel-level anomaly detection performance. As pointed out in [6, 51], in imbalanced dataset, AUROC evaluation metric might provide an inflated view of performance, which may cause difficulty in measuring

Table 7. Comparison (%) of pixel-level AUROC and AUPR of different methods on MVTec and VisA datasets.

Method	MVTec AD		VisA	
	AUROC	AUPR	AUROC	AUPR
MKD [38]	90.7	-	-	-
RDAD [7]	97.8	-	-	-
FAIR [25]	98.2	-	98.8	-
FCDD [26]	96.0	-	-	-
PatchCore [35]	**98.1**	57.6	94.7	27.8
SoftPatch [19]	97.9	-	-	-
EfficientAD [1]	96.8	60.8	98.9	**38.2**
SimpleNet [1]	97.1	51.4	91.8	22.6
HypAD (Ours)	98.0	**62.5**	**99.1**	37.6

the true capabilities of models. In anomaly detection task, the proportion of anomalous pixels is obviously lower than that of normal ones, for example, only 2.7% of all pixels in the MVTec AD test set are labeled as anomalous, and the class-imbalance problem is very prominent. In this case, we think AUPR is a more reasonable choice. For a reasonable evaluation, as shown in Tab. 7, we report the pixel-level AUROC and AUPR. From Tab. 7, we observe that AUROC score of our method is similar to the compared approaches, and pixel-level AUPR of our HypAD is better than others by 1.7% on MVTec AD and competitive to EfficientAD [1] on VisA dataset. We also can obtain from the empirical results that AUROC is considerably larger than AUPR for all methods, which is consistent with previous studies that pixel-level AUROC score exaggerates the behavior of anomaly localization methods.

5. Conclusion

We have explored the anomaly detection problem in hyperbolic space and proposed a hyperbolic anomaly detection (HypAD) method. To the best of our knowledge, this is the first attempt to investigate the anomaly detection of industrial image in hyperbolic space, breaking the limitation of the flatness of Euclidean space on the feature representations. Concretely, our proposed HypAD maps data from Euclidean space to hyperbolic space, in which the hyperbolic distance is exploited to optimize its objective. Extensive experimental results on the benchmarking datasets report that our HypAD method achieves the state-of-the-art performance, verifying its effectiveness for anomaly detection. In future work, we will make use of the existing loss functions and explore other models of hyperbolic space to learn more powerful feature representations.

Acknowledgments. This work was supported by the National Key Research and Development Program of China under Grant 2021YFB1714300.

References

[1] Kilian Batzner, Lars Heckler, and Rebecca König. Efficientad: Accurate visual anomaly detection at millisecond-level latencies. *arXiv preprint arXiv:2303.14535*, 2023. 1, 2, 4, 5, 6, 7, 8

[2] Alexander Bauer, Shinichi Nakajima, and Klaus-Robert Müller. Self-supervised training with autoencoders for visual anomaly detection. *arXiv preprint arXiv:2206.11723*, 2023. 1, 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*, pages 9592–9600, 2019. 5

[4] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 372–380, 2019. 1, 2

[5] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4183–4192, 2020. 1, 2, 5, 6

[6] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Procedings of the International Conference on Machine Learning*, pages 233–240, 2006. 8

[7] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9737–9746, 2022. 5, 6, 8

[8] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7388–7398, 2022. 8

[9] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2020. 3

[10] Aleksandr Ermolov, Leyla Mirvakhabova, Valentin Khrulkov, Nicu Sebe, and Ivan Oseledets. Hyperbolic vision transformers: Combining improvements in metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7409–7419, 2022. 2, 3, 4, 5

[11] Octavian-Eugen Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. In *Advances in Neural Information Processing Systems*, pages 5350–5360, 2018. 2, 3

[12] Zhi Gao, Yuwei Wu, Yunde Jia, and Mehrtash Harandi. Curvature generation in curved spaces for few-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8691–8700, 2021. 2

[13] Songqiao Han, Xiyang Hu, Hailiang Huang, Minqi Jiang, and Yue Zhao. Adbench: Anomaly detection benchmark. In *Advances in Neural Information Processing Systems*, pages 32142–32159, 2022. 8

[14] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 4

[15] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask r-cnn. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2961–2969, 2017. 2

[16] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. Discriminative deep metric learning for face verification in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1875–1882, 2014. 2

[17] Junlin Hu, Jiwen Lu, and Yap-Peng Tan. Sharable and individual multi-view metric learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(9):2281–2288, 2018. 2

[18] Yong Huang and Kap Luk Chan. Texture decomposition by harmonics extraction from higher order statistics. *IEEE Transactions on Image Processing*, 13(1):1–14, 2004. 1

[19] Xi Jiang, Jianlin Liu, Jinbao Wang, Qiang Nie, Kai Wu, Yong Liu, Chengjie Wang, and Feng Zheng. Softpatch: Unsupervised anomaly detection with noisy data. In *Advances in Neural Information Processing Systems*, pages 15433–15445, 2022. 8

[20] Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6418–6428, 2020. 2, 3, 4, 5

[21] Jiarui Lei, Xiaobo Hu, Yue Wang, and Dong Liu. Pyramidflow: High-resolution defect contrastive localization using pyramid normalizing flow. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14143–14152, 2023. 5, 6

[22] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9664–9674, 2021. 2

[23] Juhua Liu, Chaoyue Wang, Hai Su, Bo Du, and Dacheng Tao. Multistage gan for fabric defect detection. *IEEE Transactions on Image Processing*, 29:3388–3400, 2019. 2

[24] Shaoteng Liu, Jingjing Chen, Liangming Pan, Chong-Wah Ngo, Tat-Seng Chua, and Yu-Gang Jiang. Hyperbolic visual embedding learning for zero-shot recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9273–9281, 2020. 2, 3

[25] Tongkun Liu, Bing Li, Xiao Du, Bingke Jiang, Leqi Geng, Feiyang Wang, and Zhuo Zhao. Fair: Frequency-aware image restoration for industrial visual anomaly detection. *arXiv preprint arXiv:2309.07068*, 2023. 5, 8

[26] Philipp Liznerski, Lukas Ruff, Robert A. Vandermeulen, Billy Joe Franks, Marius Kloft, and Klaus-Robert Müller.

Explainable deep one-class classification. In *International Conference on Learning Representations*, 2021. 6, 8

[27] Takashi Matsubara, Kazuki Sato, Kenta Hama, Ryosuke Tachibana, and Kuniaki Uehara. Deep generative model using unregularized score for anomaly detection with heterogeneous complexity. *IEEE Transactions on Cybernetics*, 52 (6):5161–5173, 2020. 2

[28] Yair Movshovitz-Attias, Alexander Toshev, Thomas K Leung, Sergey Ioffe, and Saurabh Singh. No fuss distance metric learning using proxies. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 360–368, 2017. 2

[29] Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. In *Advances in Neural Information Processing systems*, pages 6338–6347, 2017. 2

[30] Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *Proceedings of the International Conference on Machine Learning*, pages 3779–3788, 2018. 2

[31] Guansong Pang, Chunhua Shen, Longbing Cao, and Anton Van Den Hengel. Deep learning for anomaly detection: A review. *ACM computing surveys (CSUR)*, 54(2):1–38, 2021. 1

[32] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. In *Advances in Neural Information Processing Systems*, pages 91–99, 2015. 2

[33] Nicolae-Cătălin Ristea, Neelu Madan, Radu Tudor Ionescu, Kamal Nasrollahi, Fahad Shahbaz Khan, Thomas B Moeslund, and Mubarak Shah. Self-supervised predictive convolutional attentive block for anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13576–13586, 2022. 1, 2

[34] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241, 2015. 2

[35] 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. 5, 6, 8

[36] 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. 2

[37] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115:211–252, 2015. 5

[38] 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*, pages 14902–14912, 2021. 1, 2, 8

[39] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015. 4

[40] Kihyuk Sohn. Improved deep metric learning with multi-class n-pair loss objective. In *Advances in Neural Information Processing Systems*, pages 1849–1857, 2016. 2

[41] Xian Tao, Xinyi Gong, Xin Zhang, Shaohua Yan, and Chandranath Adak. Deep learning for unsupervised anomaly localization in industrial images: A survey. *IEEE Transactions on Instrumentation and Measurement*, 71:1–21, 2022. 1

[42] Andreas Teuner, Olaf Pichler, and Bedrich J Hosticka. Unsupervised texture segmentation of images using tuned matched gabor filters. *IEEE Transactions on Image Processing*, 4(6):863–870, 1995. 1

[43] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in Neural Information Processing Systems*, pages 5998–6008, 2017. 4

[44] Guodong Wang, Shumin Han, Errui Ding, and Di Huang. Student-teacher feature pyramid matching for anomaly detection. In *Proceedings of the British Machine Vision Conference*, page 306, 2021. 2

[45] Xianghua Xie and Majid Mirmehdi. Texems: Texture exemplars for defect detection on random textured surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(8):1454–1464, 2007. 1

[46] Jiexi Yan, Lei Luo, Cheng Deng, and Heng Huang. Unsupervised hyperbolic metric learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12465–12474, 2021. 3

[47] Jiawei Yu, Ye Zheng, Xiang Wang, Wei Li, Yushuang Wu, Rui Zhao, and Liwei Wu. Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. *arXiv preprint arXiv:2111.07677*, 2021. 5

[48] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In *Proceedings of the British Machine Vision Conference*, 2016. 4, 5

[49] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Draema discriminatively trained reconstruction embedding for surface anomaly detection. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8330–8339, 2021. 2

[50] Hui Zhang, Zheng Wang, Zuxuan Wu, and Yu-Gang Jiang. Diffusionad: Denoising diffusion for anomaly detection. *arXiv preprint arXiv:2303.08730v2*, 2023. 5, 6, 8

[51] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *Proceedings of the European Conference on Computer Vision*, pages 392–408, 2022. 2, 4, 5, 8