

# Curved Geometric Networks for Visual Anomaly Recognition

Jie Hong<sup>1</sup>, Pengfei Fang<sup>1</sup>, Weihao Li, Junlin Han, Lars Petersson, and Mehrtash Harandi<sup>2</sup>

**Abstract**—Learning a latent embedding to understand the underlying nature of data distribution is often formulated in Euclidean spaces with zero curvature. However, the success of the geometry constraints, posed in the embedding space, indicates that curved spaces might encode more structural information, leading to better discriminative power and hence richer representations. In this work, we investigate the benefits of the curved space for analyzing anomalous, open-set, or out-of-distribution (OOD) objects in data. This is achieved by considering embeddings via three geometry constraints, namely, spherical geometry (with positive curvature), hyperbolic geometry (with negative curvature), or mixed geometry (with both positive and negative curvatures). Three geometric constraints can be chosen interchangeably in a unified design, given the task at hand. Tailored for the embeddings in the curved space, we also formulate functions to compute the anomaly score. Two types of geometric modules (i.e., geometric-in-one (GiO) and geometric-in-two (GiT) models) are proposed to plug in the original Euclidean classifier, and anomaly scores are computed from the curved embeddings. We evaluate the resulting designs under a diverse set of visual recognition scenarios, including image detection (multiclass OOD detection and one-class anomaly detection) and segmentation (multiclass anomaly segmentation and one-class anomaly segmentation). The empirical results show the effectiveness of our proposal through consistent improvement over various scenarios. The code is made available at <https://github.com/JHome1/GiO-GiT>.

**Index Terms**—Anomaly recognition, geometric learning, hyperbolic space, mixed-curvature space, open-set recognition, out-of-distribution (OOD) detection, spherical space.

## I. INTRODUCTION

IN THIS article, we aim to leverage the curved geometry for learning embeddings, which in return allows us to analyze and identify anomalous, open-set, or out-of-distribution (OOD) objects from normal, closed-set, or in-distribution (ID) input data. Nonflat geometry has gained an increasing amount of interest in various machine learning approaches due to its intriguing properties in encoding the inherent geometry or hidden structure information of the data [1], [2], [3], [4], [5], [6]. For example, spherical spaces with constant positive curvature show appealing properties along with deep neural networks (DNNs) to encode samples resembling the sphere [7], [8]. Hyperbolic spaces, featured with a constant negative curvature, are shown to be rich in encoding the underlying hierarchical structure in the data. Such a property enables hyperbolic spaces to better discriminate input samples [4]. Fig. 1 conceptualizes the distinctive characteristics of Euclidean, spherical, and hyperbolic spaces and highlights their disparities. Euclidean spaces with zero curvature are familiar faces in DNNs [see Fig. 1(a)]. Spherical geometry, as shown in Fig. 1(b), has been successfully employed to encode directional data (i.e., samples where the magnitude does not carry important information). The Poincaré ball model for the hyperbolic spaces, as shown in Fig. 1(c), provides a structure with constant negative curvature to encode data.

Previous studies, such as [4], [11], and [12], show that curved spaces can attain a superior performance gain over the Euclidean space, especially for tasks relying on image embeddings (e.g., zero-/few-shot learning or metric learning). For example, in [12], by employing the similarity metric in spherical embedding spaces, the model enhances its discriminative ability in zero-shot classification for unknown classes. In [4] and [13], hyperbolic spaces are shown to have a better distribution across unknown classes, therefore improving few-shot learning performance. Based on the interpretations in Fig. 1, we infer that the curved geometric embeddings exhibit enhanced discrimination due to the following reasons.

- 1) In the spherical space, all points are constrained to lie on the surface of a hypersphere, as shown in Fig. 1(b). On the surface of the hypersphere, the embedding

Manuscript received 28 July 2022; revised 27 May 2023; accepted 18 August 2023. This work was supported in part by the National Science Foundation of China under Grant 62306070 and in part by the Southeast University Start-Up Grant for New Faculty under Grant 4009002309. The work of Mehrtash Harandi was supported by the Australian Research Council (ARC) under Project DP230101176. (Corresponding author: Pengfei Fang.)

Jie Hong and Junlin Han are with the College of Engineering, Computing and Cybernetics, Australian National University, Canberra, ACT 2601, Australia, and also with Data61-CSIRO, Black Mountain Laboratories, Canberra, ACT 2601, Australia (e-mail: jie.hong@anu.edu.au; u6835134@anu.edu.au).

Pengfei Fang is with the School of Computer Science and Engineering and the Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications, Ministry of Education, Southeast University, Nanjing, 210096, China (e-mail: fangpengfei@seu.edu.cn).

Weihao Li is with the College of Engineering, Computing and Cybernetics, Australian National University, Canberra, ACT 2601, Australia (e-mail: weihao.li1@anu.edu.au).

Lars Petersson is with Data61-CSIRO, Black Mountain Laboratories, Canberra, ACT 2601, Australia (e-mail: lars.petersson@data61.csiro.au).

Mehrtash Harandi is with the Department of Electrical and Computer Systems Engineering, Monash University, Clayton, VIC 3800, Australia (e-mail: mehrtash.harandi@monash.edu).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TNNLS.2023.3309846>, provided by the authors.

Digital Object Identifier 10.1109/TNNLS.2023.3309846

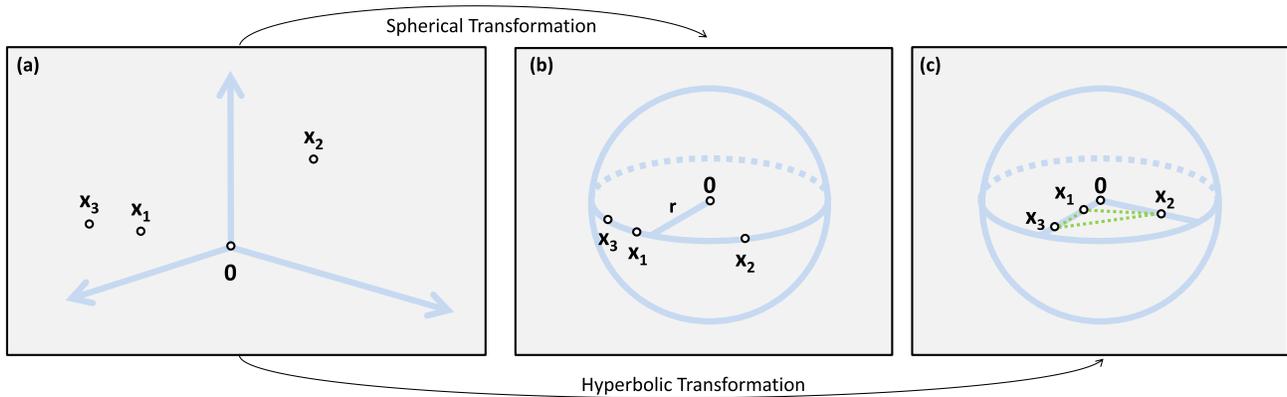


Fig. 1. Illustration of Euclidean space and curved spaces (i.e., spherical and hyperbolic spaces). (a) Euclidean space represents a flat space without curvature. (b) In contrast, the spherical space restricts all points to lie on the surface of a hypersphere. (c) Hyperbolic space, represented by the Poincaré ball model, makes points exist inside the ball, demonstrating its negatively curved nature. The transformation from Euclidean space to spherical or hyperbolic space can be achieved through respective spherical or hyperbolic transformation, which project points onto the corresponding curved space.

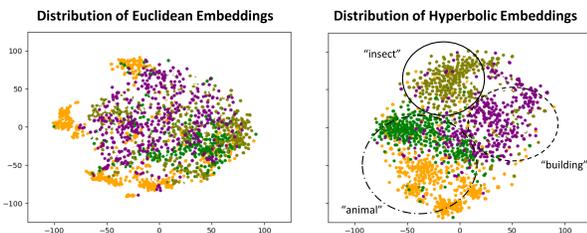


Fig. 2. Illustration of the discrimination ability of curved spaces on unknown data. We train two models using Euclidean and hyperbolic geometry on *miniImageNet* dataset [9], following the few-shot learning protocol. We randomly choose four unknown classes and plot the class embeddings using t-SNE [10]. Each unknown class is represented by one color. From this visualization, we can directly observe that within each unknown class, embeddings in hyperbolic space (see the figure on the right) show a more compact distribution than embeddings in Euclidean space (see the figure on left). Moreover, it can be observed that hyperbolic embeddings exhibit improved separations based on the discovered hierarchies. For instance, when compared to the distribution of Euclidean embeddings, hyperbolic embeddings show closer proximity between classes (i.e., the yellow-colored “African hunting dog” and the green-colored “lion”) that belong to the same superclass (i.e., “animal”). Conversely, hyperbolic embeddings of classes (i.e., the olive-colored “ant” and the purple-colored “bookshop”), which belong to different superclasses (i.e., “insect” and “building”), are pulled further apart.

space is enforced to have a uniform distribution across multiple classes, promoting a more unbiased classification. Furthermore, scale-free analysis, the inclusion of spherical spaces, e.g., via  $\ell_2$  normalization in a DNN, provides useful constraints for learning and has been used with success in the majority of modern neural structures [7], [8], [14].

- As shown in Fig. 1(c), points in the hyperbolic space lie inside the ball. The volume in a hyperbolic space expands exponentially with dimensions. This enables us to embed tree-like structures with less distortion (see the green-colored triangle) and aids the discovery of hierarchical relationships among samples. For instance, as shown in Fig. 2 and compared to Euclidean embeddings, the hyperbolic embeddings of “African hunting dog” (yellow) and “lion” (green) are closer, emphasizing their similarity at the superclass level (i.e., “animals”). Moreover, the hyperbolic embeddings of

“ant” (olive) and “bookshop” (purple) exhibit better separation than Euclidean embeddings as they originate from distinct superclasses, “insect” and “building.” Learning such hierarchies will lead to success at the inference.

With the appealing observations shown above, in this article, we investigate the practice of using geometries with fixed nonzero curvature in visual anomaly, open-set, or OOD recognition tasks. For the purpose of realizing our idea, a natural solution is to simply replace the existing Euclidean classifier with one based on a curved geometry. This idea is behind the design of our geometric-in-one (GiO) model. In addition, we also find that the “divergence” between Euclidean embeddings and curved embeddings can provide a reliable indicator useful for anomaly, open-set, or OOD identification. To benefit from this interesting observation, we further develop a geometric-in-two (GiT) model. Having multiple geometry-aware networks at our disposal, we further present approaches for getting the anomaly score to identify abnormal objects.

Our main objective is to show that geometry plays an essential role in identifying anomalies. As such, we develop a generic solution and incorporate it into various baselines in our empirical study. The contributions of this work can be summarized as follows.

- We propose two types of lightweight curvature-aware geometric networks for visual anomaly, open-set, or OOD recognition. To the best of our knowledge, this is the first attempt to adopt curved manifolds as embedding spaces to distinguish normal/closed-set/ID and anomalous/open-set/OOD data. In addition, multiple curved spaces, including spherical, hyperbolic, and mixed spaces, are studied.
- Extensive experiments on a wide range of visual anomaly, open-set, or OOD recognition tasks (e.g., multiclass OOD detection, one-class anomaly detection, multiclass anomaly segmentation, and one-class anomaly segmentation) suggest that the proposed technique leads to a substantial performance gain over the Euclidean geometry.

## II. RELATED WORK

### A. Visual Anomaly Recognition

Three main approaches are developed for doing visual anomaly (or open-set/OOD) recognition: confidence-, generative-, and self-supervised-based methods. However, a few works are learning features in non-Euclidean spaces.

1) *Confidence-Based Method*: It is well known that the confidence from the softmax in a classifier helps to detect OOD samples from ID samples since ID samples are more likely to have a greater maximum softmax confidence compared to OOD samples [15]. OOD detector for neural network (ODIN) [16] applies temperature scaling to the confidence vector and adds small perturbations to input samples for more accurate OOD detection. Additional confidence-based methods, which make use of the confidence, have been studied in [17], [18], [19], [20], and [21].

2) *Generative-Based Method*: One of the generative-based methods is to synthesize effective training samples to avoid the DNNs becoming overconfident in their predictions [22], [23], [24], [25], [26]. Another choice is to optimize features in the latent space of an encoder–decoder network toward generating a more general distribution [27], [28], [29], [30], [31] or a more representative attention map [32], [33], [34].

3) *Self-Supervised-Based Method*: Self-supervised learning techniques have been widely employed in anomaly recognition. Ensemble leave-out classifier (ELOC) [35] trains classifiers in a self-supervised manner by setting a subset of training data as OOD data. One main idea behind the self-learning method is to apply geometric transformations (GTs) or augmentations on the input images and train a multiclass model to discriminate such transformations (or augmentations). Prediction of image rotation is used in rotation network (RotNet) [36]. Jittered patches of an image are classified in Patch-SVDD [37] for anomaly localization. Another idea is to use contrastive learning for better visual representations [38], [39], [40]. More works using self-supervised learning are presented in [38], [41], [42], [43], [44], [45], and [46].

Some other facts should be noted.

- 1) Confidence-based as well as self-supervised-based methods mainly adopt “encoder–classifier” structures and generative-based models with “encoder–decoder” architectures. The proposed modules in our work are best applied to an “encoder–classifier” rather than an “encoder–decoder” structure.
- 2) Teacher–student structures have been utilized in anomaly recognition problems [47], [48], [49], [50]. These approaches demonstrate the effectiveness of leveraging the discrepancy between the teacher and student branches to identify and localize anomalies. In [49], a pair of networks are employed for anomaly detection, where the teacher network is pretrained on ImageNet. The discrepancy between multilevel layers is utilized to compute anomaly scores. In the case of the local–global net proposed in [50], a teacher–student architecture is trained to compute local and global features. In the

inference, local and global features are compared to compute the anomaly score.

- 3) Some approaches compute the anomaly score of test samples by comparing their features with those of training samples [37], [45]. In the Patch-SVDD method [37], for instance, the anomaly scores of test samples are obtained by calculating the  $L_2$  distance between the features of the test and training samples.

### B. Geometric Learning

Geometric learning has been studied extensively to encode structured representations [51]. For example, the set has been used to model order-invariant data (e.g., 3-D point clouds [52] or video data [53]). Orthogonal constraints, i.e., subspaces, are often used to encode set data [54], [55], for its potential to be robust against illumination variations, background, and so on. In vectorized representations, spherical or hyperbolic spaces are also very effective for metric learning-related tasks. In the spherical space, the similarity of representations is upper bounded. Hence, such a space is particularly well-behaved at learning a metric space [7], [56], [57]. As opposed to the spherical space, tree-like data can be embedded in the hyperbolic space for its intriguing property to capture the hierarchical structure of the data [4], [11], [58]. To further increase the discrimination power of the learned embeddings in curved spaces, the kernel methods, which implicitly map the geometric representation to a high or even infinite-dimensional feature space, are studied for spherical embeddings [59] or hyperbolic embeddings [13]. To fully model the structure of the data, mixed-curved spaces are good candidates as embedding spaces [14], [60].

## III. PRELIMINARIES AND BACKGROUND

In this section, we will briefly introduce the preliminary knowledge and background used in this article.

### A. Notation

We use  $\kappa$  to denote the curvature of a manifold. In general, a vectorized representation or an embedding can be embedded in three types of manifolds: the Euclidean space  $\mathcal{M}_E$ , the spherical space  $\mathcal{M}_S$ , and the hyperbolic space  $\mathcal{M}_H$ , corresponding to  $\kappa = 0$ ,  $\kappa > 0$ , and  $\kappa < 0$ , respectively. Throughout this article, we call any space with  $\kappa \neq 0$ , as a curvature-aware space or a curved space. A mixed-curvature manifold  $\mathcal{M}_M$  is a product space, consisting of a set of different spaces [14], [60]. In our work, the mixed-curvature manifold is defined as  $\mathcal{M}_M = \mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{M}_3 \times \cdots \times \mathcal{M}_N$ , in which we mix  $N$  different manifolds. For example, the mixed-curvature manifold  $\mathcal{M}_M = \mathcal{M}_E \times \mathcal{M}_S \times \mathcal{M}_H$  includes a Euclidean space, a spherical space, and a hyperbolic space.

### B. Spherical Geometry

The  $n$ -sphere with curvature  $\kappa > 0$  is defined as

$$\mathbb{S}_\kappa^{n-1} = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\|^2 = 1/\kappa\}. \quad (1)$$

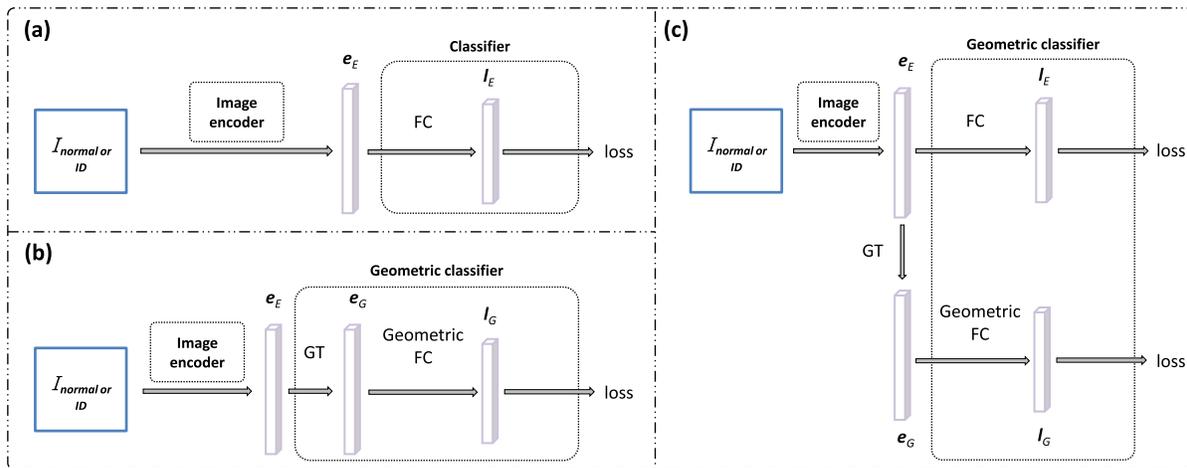


Fig. 3. General frameworks of the models and their training processes are depicted. (a) Baseline model follows an “encoder–classifier” structure. The image encoder takes the input image  $I$  and encodes it into an embedding vector  $e_E$ . The classifier, primarily composed of an FC layer, generates the confidence vector for loss computation. We propose two models in our work: GiO and GiT. GiO and GiT differ from the baseline model in that they integrate the GT and geometric FC layer into the classifier. In essence, while the baseline model follows an “encoder–classifier” structure, GiO and GiT adopt an “encoder-geometric-classifier” structure. (b) In the GiO model, the original Euclidean FC layer is replaced with a geometric FC layer based on a curved geometry. (c) GiT model adopts the geometric FC layer as the additional branch. In the training stage, we follow the setting that the training data do not include any anomalous or OOD samples.

The mapping  $\Gamma_S : \mathbb{R}^n \rightarrow \mathbb{S}_\kappa^{n-1}$  projects an embedding  $\mathbf{x} \in \mathbb{R}^n$  generated by an image encoder to  $n$ -sphere as

$$\mathbf{x}_S = \Gamma_S(\mathbf{x}) = \frac{\mathbf{x}}{\sqrt{\kappa \|\mathbf{x}\|}}. \quad (2)$$

The embedding  $\mathbf{x}_S$  indicates a point in the  $n$ -sphere, satisfying the constraint in (1). In practice, the angular mapping in the  $n$ -sphere, analogous to the linear mapping in the Euclidean space, can be realized by a fully connected (FC) layer with weight  $\mathbf{W}$ . Let  $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_j, \dots, \mathbf{w}_C]$ , where  $\mathbf{w}_j \in \mathbb{S}_\kappa^{n-1}$  is the corresponding column, representing the parameters of the classifier. The prediction associated with the  $j$ th class for an embedding  $\mathbf{x}$  is determined by  $\mathbf{w}_j$ . We note that for  $\mathbf{x}, \mathbf{w}_j \in \mathbb{S}_\kappa^{n-1}$ , the term for the  $j$ th class,  $l_S(\mathbf{x}, \mathbf{w}_j) = \langle \mathbf{x}, \mathbf{w}_j \rangle$ , is indeed related to the geodesic distance on  $\mathbb{S}_\kappa^{n-1}$ ; hence, one can understand this term as a form of the distance-based value. Here, we use the notation  $l_S(\mathbf{x}, \mathbf{W})$  to show a vector obtained by applying the columns of  $\mathbf{W}$  to  $\mathbf{x} \in \mathbb{S}_\kappa^{n-1}$ .

For the spherical network, we compute the angular loss  $\ell_S$  based on  $B$  samples in one batch

$$\ell_S = -\frac{1}{B} \sum_i \log \frac{\exp(l_{S,y_i})}{\sum_j \exp(l_{S,j})} \quad (3)$$

where  $l_{S,j}$  is the  $j$ th element in  $l_S(\mathbf{x}, \mathbf{W})$  under the  $i$ th input sample and  $\mathbf{x} \in \mathbb{S}_\kappa^{n-1}$ . Accordingly,  $l_{S,y_i}$  is the  $y_i$ th element in  $l_S(\mathbf{x}, \mathbf{W})$  and  $y_i$  indicates the label class to the  $i$ th input sample.

### C. Hyperbolic Geometry

In contrast to the  $n$ -sphere  $\mathbb{S}_\kappa^{n-1}$ , the hyperbolic space is a curved space with a constant negative curvature (i.e.,  $\kappa < 0$ ). The hyperbolic space offers an appealing property for anomaly problems. Specifically, the volume in the hyperbolic space increases exponentially. This allows the algorithm to incorporate more embeddings of unknown objects, particularly

those closer to the origin, as the space expands from the origin [4].

In this article, we employ the Poincaré ball [1], [4] to model and work with the hyperbolic space. The  $n$ -dimensional Poincaré ball, with curvature  $\kappa$ , is defined by the manifold

$$\mathbb{H}_\kappa^n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < -1/\kappa\}^1$$

To embed  $\mathbf{x} \in \mathbb{R}^n$ , obtained by an image encoder to the Poincaré ball, we use the following transformation:

$$\mathbf{x}_H = \Gamma_H(\mathbf{x}) = \begin{cases} \mathbf{x}, & \text{if } \|\mathbf{x}\| \leq \frac{1}{|\kappa|} \\ \frac{1 - \xi}{|\kappa|} \frac{\mathbf{x}}{\|\mathbf{x}\|}, & \text{else} \end{cases} \quad (4)$$

where  $\xi > 0$  is a small value to ensure numerical stability. The embedding  $\mathbf{x}_H$  is a point in the Poincaré ball. To enable the vector operations in the Poincaré ball, we make use of the Möbius addition for  $\mathbf{x}, \mathbf{y} \in \mathbb{H}_\kappa^n$  as

$$\mathbf{x} \oplus_\kappa \mathbf{y} = \frac{(1 + 2|\kappa| \langle \mathbf{x}, \mathbf{y} \rangle + |\kappa| \|\mathbf{y}\|^2) \mathbf{x} + (1 - |\kappa| \|\mathbf{x}\|^2) \mathbf{y}}{1 + 2|\kappa| \langle \mathbf{x}, \mathbf{y} \rangle + |\kappa|^2 \|\mathbf{x}\|^2 \|\mathbf{y}\|^2} \quad (5)$$

where  $\langle \cdot, \cdot \rangle$  is the inner product. The geodesic distance between  $\mathbf{x}, \mathbf{y} \in \mathbb{H}_\kappa^n$  is defined as

$$d_{\text{Geo}}(\mathbf{x}, \mathbf{y}) = \frac{2}{\sqrt{|\kappa|}} \tanh^{-1} \left( \sqrt{|\kappa|} \|\mathbf{x} \oplus_\kappa \mathbf{y}\| \right). \quad (6)$$

One can also generalize the hyperbolic linear operation, parameterized by  $\mathbf{W}$  (e.g., the hyperbolic linear layer), in the Poincaré ball [4]

$$\mathbf{W} \oplus_\kappa(\mathbf{x}) := \frac{1}{\sqrt{|\kappa|}} \tanh \left( \frac{\|\mathbf{W}\mathbf{x}\|}{\|\mathbf{x}\|} \tanh^{-1} \left( \sqrt{|\kappa|} \|\mathbf{x}\| \right) \right) \frac{\mathbf{W}\mathbf{x}}{\|\mathbf{W}\mathbf{x}\|}. \quad (7)$$

The proposed network contains the multiclass classification layer. We employ the generalization of multiclass logistic regression (MLR) to the hyperbolic spaces [4]. Following the

<sup>1</sup>In this case, the Riemannian metric is defined as  $g_\kappa^{\mathbb{H}}(\mathbf{x}) = \lambda_\kappa^2(\mathbf{x}) \cdot g^E$ , where  $\lambda_\kappa(\mathbf{x}) = \frac{1}{1 + \kappa \|\mathbf{x}\|^2}$ , and  $g^E$  is the Euclidean metric.

work in [4], the formulation of the hyperbolic MLR for  $C$  classes is given by:

$$l_H(y = j|\mathbf{x}) \propto \exp\left(\frac{\lambda^\kappa(\mathbf{x})\|\mathbf{W}_j\|}{\sqrt{|\kappa|}} \sinh^{-1}\left(\frac{2\sqrt{|\kappa|}\langle -\mathbf{p}_j \oplus_\kappa \mathbf{x}, \mathbf{W}_j \rangle}{(1 - |\kappa| \cdot \|\mathbf{p}_j \oplus_\kappa \mathbf{x}\|^2)\|\mathbf{W}_j\|}\right)\right) \quad (8)$$

where  $j \in \{1, 2, \dots, C\}$ . Here,  $\mathbf{x} \in \mathbb{H}_\kappa^n$  is an embedding in the hyperbolic space, and  $\mathbf{p}_j \in \mathbb{H}_\kappa^n$ ,  $\mathbf{W}_j \in T_{\mathbf{p}_j}\mathbb{H}_\kappa^n \setminus \{\mathbf{0}\}$  are learnable weights.

#### IV. APPROACH

Visual anomaly (or open-set/OOD) recognition aims to identify abnormal (or open-set/OOD) samples from normal (or closed-set/ID) samples. During the training process, as shown in Fig. 3, only normal, closed-set, or ID data can be accessed. For the evaluation stage, as shown in Fig. 4, both normal (or ID) and anomalous (or OOD) inputs to be recognized exist.

The pipeline of the baseline is shown in Fig. 3(a) for the training phase and Fig. 4(a) for the inference phase. Specifically, an image encoder first maps the input image to a feature embedding  $\mathbf{e}_E$ , lying in a Euclidean space. The following FC layer and Softmax function further predict the probability belonging to each normal (or ID) class, denoted by  $\mathbf{c}_E$  [see Fig. 3(a)]. In the evaluation process, to identify whether the input image or pixel is an outlier (also known as anomalous, open-set, or OOD data), one needs to define the anomaly score  $AS \in [0, 1]$ . In this vanilla model [see Fig. 4(a)], we follow the common practice in [15] to define the anomaly score by leveraging the predication  $\mathbf{c}_E$  as  $AS = 1 - \max(\mathbf{c}_E)$ .

A curvature-aware geometric model indicates a model where its classifier operates in a curved space  $\mathcal{M}_G$ , and we term its classifier geometric classifier. Two curvature-aware geometric models are presented in this section: GiO and GiT, as shown in Fig. 3(b) and (c). Compared to the baseline model, the curvature-aware geometric model has two geometric layers, namely, GT and geometric FC layer. The GT is to transform the Euclidean embedding  $\mathbf{e}_E$  computed by an image encoder to the geometric embedding  $\mathbf{e}_G$  (see (2) for the spherical geometry  $\mathcal{M}_S$  and (4) for the hyperbolic geometry  $\mathcal{M}_H$ ). The geometric FC layer is a generalization of the FC layer in  $\mathcal{M}_G$  (e.g., the angular linear layers in  $\mathcal{M}_S$  or the hyperbolic linear layers in  $\mathcal{M}_H$ ). As shown in Fig. 3, we learn geometric classifiers where embeddings extracted from an image encoder are manipulated in the curved spaces. During the inference phase, as shown in Fig. 4(b) and (c), the geometric score  $z$  is initially obtained from the curved embedding. In GiO, the variable  $z$  can be interpreted as the distance value between the curved embedding point and the reference point in the curved space. In GiT,  $z$  represents the discrepancy between two embeddings originating from different spaces. Subsequently, the anomaly score in our geometric model is computed as  $AS = 1 - \tanh(z)$ . Since  $z \in [0, +\infty)$  is not constrained within the range of 0–1, the

function  $\tanh(\cdot)$  is employed to make sure that the anomaly score AS falls between 0 and 1.

##### A. GiO Model

We first introduce a single-branch framework, i.e., the GiO model. The GiO model [see Fig. 3(b)] is a natural modification that replaces the Euclidean embeddings with curved geometry embeddings. It provides a straightforward comparison to justify the advantages of using curved geometry for anomaly recognition. Specifically, as shown in Fig. 3(b), the image or pixel embeddings can be obtained by applying the GT function (i.e., (2) for  $\mathcal{M}_S$  and (4) for  $\mathcal{M}_H$ ) to the feature vectors encoded by the image encoder. Then, the following geometric classifier, realized by the geometric FC layer, is further used to predict the class of its input. For example, the hyperbolic-in-one (HiO) model refers to a network with a geometric classifier in the hyperbolic space. Compared to the baseline model [see Fig. 3(a)], the GiO model only modifies the embedding layer and classifier without bringing extra parameters, thereby being a cheap yet flexible solution.

In the GiO model, the geometric score  $z_G \in \{z_S, z_H, z_M\}$  can be obtained from embedding  $\mathbf{e}_G$ . In our work, the geometric score from the spherical manifold  $\mathcal{M}_S$  is defined as

$$z_S(\mathbf{e}_S) = \max(l_S(\mathbf{e}_S, \mathbf{W})) \quad (9)$$

where  $\mathbf{e}_S \in \mathbb{S}_\kappa^{n-1}$ . Experiments in [7] and [31] verified that  $z_S$  from  $\mathcal{M}_S$  is suitable for visual tasks under the open-set protocol. Hence, we expect that  $z_S$  could help in anomaly, open-set, or OOD recognition. For the hyperbolic space  $\mathcal{M}_H$ , the geometric score is defined as

$$z_H(\mathbf{e}_H) = d_{\text{Geo}}(\mathbf{e}_H, \mathbf{0}_H) \quad (10)$$

where  $d_{\text{Geo}}(\mathbf{e}_H, \mathbf{0}_H)$  is also known as the geodesic distance between the point  $\mathbf{e}_H$  and the origin  $\mathbf{0}_H$ , for  $\mathbf{e}_H, \mathbf{0}_H \in \mathbb{H}_\kappa^n$ . The experiment of image-level OOD detection in [4] shows the property of  $\mathbf{e}_H$ , and the hyperbolic embedding points of OOD samples are closer to the origin. In the mixed-curvature manifold  $\mathcal{M}_M$ , the geometric score  $z_M$  is formulated as

$$z_M^2 = \sum_i^N z_{M,i}^2 \quad (11)$$

where there are  $N$  spaces in  $\mathcal{M}_M$  and  $z_{M,i}$  is the geometric score from the  $i$ th component space. For instance, in  $\mathcal{M}_M = \mathcal{M}_S \times \mathcal{M}_H$ ,  $z_M^2 = z_S^2 + z_H^2$ . The GiO model with  $\mathcal{M}_M = \mathcal{M}_S \times \mathcal{M}_H$  is actually a two-branch architecture. We can even incorporate the Euclidean space into the mixed-in-one (MiO) model. In  $\mathcal{M}_M = \mathcal{M}_E \times \mathcal{M}_S \times \mathcal{M}_H$ , we can have  $z_M^2 = z_E^2 + z_S^2 + z_H^2$  where the geometric score from Euclidean space is  $z_E = \max(\mathbf{c}_E)$  [15]. Table I lists the proposed geometric networks, in conjunction with the geometry score, in the GiO model.

Having the geometry score  $z_G$  at our disposal, the anomaly score is defined as  $AS = 1 - \tanh(z_G)$ . A higher value of AS indicates a higher probability that the input is coming from the anomalous distribution. The purpose of  $\tanh(\cdot)$  is to normalize  $z_G$  to a value that is in the range of 0–1 [see Fig. 4(b)].

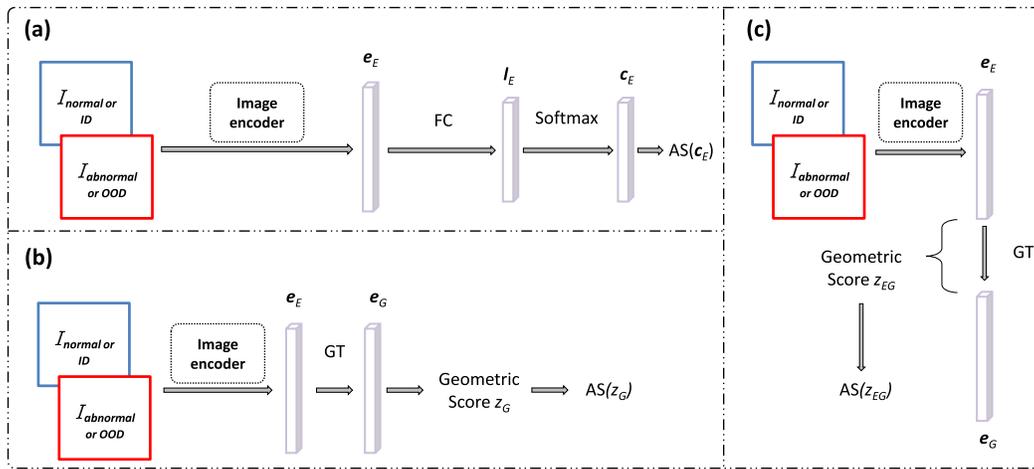


Fig. 4. General frameworks of the models and their evaluation processes are depicted. In the evaluation stage, the model would process both normal (or ID) and anomalous (or OOD) inputs. The geometric score  $z_G$  is extracted from the curved embedding for anomaly score computation in the GiO model. For the GiT model, the geometric score  $z_{EG}$  is obtained via Euclidean and curved embeddings. During the evaluation process, GiO and GiT models use the less parameters than the baseline. (a) Baseline model. (b) GiO model. (c) GiT model.

TABLE I

PROPOSED GEOMETRIC NETWORKS IN THE GiO MODEL AND THE GiT MODEL. “ABB.,” “GEO. COMPONENTS,” AND “GEO. SCORE” INDICATE “ABBREVIATION,” “GEOMETRY COMPONENTS,” AND “GEOMETRY SCORE,” RESPECTIVELY

Method	Abb.	Geo. Components	Geo. Score
Geometric-in-One (GiO) Model			
Spherical-in-One	SiO	$\mathcal{M}_S$	Eq. (9)
Hyperbolic-in-One	HiO	$\mathcal{M}_H$	Eq. (10)
Mixed-in-One	MiO	Multiple $\mathcal{M}$	Eq. (11)
Geometric-in-Two (GiT) Model			
Spherical-in-Two	SiT	$\mathcal{M}_S$	Eq. (12)
Hyperbolic-in-Two	HiT	$\mathcal{M}_H$	Eq. (12)
Mixed-in-Two	MiT	Multiple $\mathcal{M}$	Eq. (13)

### B. GiT Model

Several recent works [47], [48], [49], [50] develop the dual-branch architecture and exploit the discrepancy between features in separate classifiers for anomaly, open-set, or OOD recognition. Motivated by this fact, we further introduce our second framework, termed GiT, where a Euclidean classifier and a geometric classifier are integrated after the image encoder [see Fig. 3(c)]. In the GiT model, in parallel with a branch of the Euclidean classifier, the other branch learns the feature embedding in the curved space, and the following geometric FC layer is used as a class predictor. The embedding in the curved space  $\mathcal{M}_G$  is achieved by transforming  $\mathbf{e}_E$  to  $\mathbf{e}_G$  via a GT function, as shown in Fig. 3(c). In such a pipeline, the geometry-aware score  $z_{EG}$  is defined as the discrepancy between distributions of  $\mathbf{e}_E$  and  $\mathbf{e}_G$ , measured via the Kullback–Leibler (KL) divergence, as follows:

$$z_{EG} = \sum_i p_{E,i} \log \frac{p_{E,i}}{p_{G,i}} \quad (12)$$

where  $p_{E,i}$  and  $p_{G,i}$  are the  $i$ th element in  $\mathbf{p}_E$  and  $\mathbf{p}_G$ , respectively. Here,  $\mathbf{p}_E = \text{softmax}(\mathbf{e}_E)$  and  $\mathbf{p}_G = \text{softmax}(\mathbf{e}_G)$ . The geometric score  $z_{EG}$  is essentially the distribution discrepancy between the learned embedding  $\mathbf{e}_E$  from  $\mathcal{M}_E$  and  $\mathbf{e}_G$  from  $\mathcal{M}_G$ . We have three types of GiT models:

spherical-in-two (SiT), hyperbolic-in-two (HiT), and mixed-in-two (MiT) models, and thereby,  $z_{EG} \in \{z_{ES}, z_{EH}, z_{EM}\}$  where  $z_{ES}$ ,  $z_{EH}$ , and  $z_{EM}$  are from three models. The values  $z_{ES}$  and  $z_{EH}$  can be easily calculated by (12). Inspired by (11), we define  $z_{EM}$  in  $\mathcal{M}_M$  as follows:

$$z_{EM}^2 = \sum_i^N z_{EM,i}^2 \quad (13)$$

where  $N$  indicates the number of component spaces in  $\mathcal{M}_M$ . For example, when  $\mathcal{M}_M = \mathcal{M}_S \times \mathcal{M}_H$ , the score metric can be obtained by  $z_{EM}^2 = z_{ES}^2 + z_{EH}^2$ . The GiT model with  $\mathcal{M}_M = \mathcal{M}_S \times \mathcal{M}_H$  is actually a three-branch architecture.

Our experiments show that  $z_{EG}$  is able to provide reliable discrimination information for anomaly identification. We find that similar to our GiO models, the anomaly score of GiT models should be  $\text{AS} = 1 - \tanh(z_{EG})$  [see Fig. 4(c)]. Table I lists the networks and the geometry score in the GiT model.

### C. Model Training

In the baseline model, as shown in Fig. 3(a), the Euclidean classifier is optimized by a standard cross-entropy loss function, as  $\ell = \ell_E(\mathbf{c}_E)$ . Similarly, we optimize the GiO model using the confidence vector  $\mathbf{c}_G$ , predicted in its geometric classifier with its own specific loss  $\ell = \ell_G(\mathbf{c}_G)$  [see Fig. 3(b)]. The loss functions for the spherical and hyperbolic geometric networks are described in (3) and (8).

The GiT model, as shown in Fig. 3(c), is trained in a multitask learning manner by optimizing a Euclidean classifier and a geometric classifier, as  $\ell = \ell_E(\mathbf{c}_E) + \ell_G(\mathbf{c}_G)$ . To be more specific, a shared image encoder encodes the input image in a Euclidean space  $\mathcal{M}_E$  and the curved spaces  $\mathcal{M}_G$ . Then, the following Euclidean classifier and geometric classifier are optimized separately. In contrast to the well-studied student-teacher models [48], [49], which aim to transfer the knowledge from the teacher model to the student model, our GiT model learns subbranches guided by its own spaces and objective functions [see (3) and (8)].

TABLE II  
VISUAL ANOMALY TASKS WHERE WE EVALUATE THE PROPOSED GEOMETRIC MODELS

Task	Recognition Space	Num. normal or ID Class	Num. abnormal or OOD Class
Multi-Class OOD Detection	Image	Multiple	Multiple
One-Class Anomaly Detection	Image	One	Multiple
Multi-Class Anomaly Segmentation	Pixel	Multiple	Multiple
One-Class Anomaly Segmentation	Pixel	One	Multiple

TABLE III

MULTICLASS OOD DETECTION ON CIFAR-10/CIFAR-100 [61] WITH DENSENET/WRN-28-10 IMAGE ENCODER. MIXED-GEOMETRY EMBEDDING IN THE MiO/MiT MODEL INCLUDES SPHERICAL EMBEDDING AND HYPERBOLIC EMBEDDING. WE PROVIDE AVERAGING RESULTS OVER FIVE MULTIPLE OOD DATASETS: TINc, TINr, LSUNc, LSUNr, AND iSUN. “\*” INDICATES THAT THE RESULTS ARE OBTAINED VIA A SELF-IMPLEMENTED NETWORK. THE CURVATURES OF SiO, SiT, HiO, HiT, MiO, AND MiT ARE SET TO 1.0, 1.0,  $-0.01$ ,  $-1.0$ ,  $(1.0, -0.01)$ , AND  $(1.0, -1.0)$ , RESPECTIVELY

Image encoder (on ID dataset)	Method	FPR (95% TPR) ↓	Detection Error ↓	AUROC ↑
Dense-BC (on CIFAR-10)	Hendrycks&Gimpel* [15]	36.9	10.9	94.8
	SiO/HiO/MiO	37.8/6.3/8.3	11.0/5.4/6.2	94.7/98.7/98.5
	SiT/HiT/MiT	10.5/20.8/11.2	6.7/10.0/7.0	98.1/96.0/98.0
Dense-BC (on CIFAR-100)	Hendrycks&Gimpel* [15]	70.7	26.3	81.5
	SiO/HiO/MiO	69.4/43.0/52.4	26.1/17.1/21.5	81.4/90.7/86.8
	SiT/HiT/MiT	52.6/55.9/66.4	20.3/21.8/25.7	87.5/85.9/81.7
WRN-28-10 (on CIFAR-10)	Hendrycks&Gimpel* [15]	41.9	12.5	93.5
	SiO/HiO/MiO	43.7/16.4/17.9	12.7/8.5/9.3	93.4/97.1/96.9
	SiT/HiT/MiT	15.5/28.7/20.1	8.1/12.5/9.8	97.1/94.1/96.3
WRN-28-10 (on CIFAR-100)	Hendrycks&Gimpel* [15]	70.8	26.4	81.5
	SiO/HiO/MiO	69.9/51.1/49.2	26.3/20.3/19.2	81.5/88.1/88.8
	SiT/HiT/MiT	49.3/56.4/57.1	20.0/21.7/21.8	88.2/85.7/86.4

TABLE IV

MULTICLASS OOD DETECTION ON CIFAR-10/CIFAR-100 [61] WITH DENSENET/WRN-28-10 IMAGE ENCODER. WE PROVIDE AVERAGING RESULTS OVER FIVE MULTIPLE OOD DATASETS: TINc, TINr, LSUNc, LSUNr, AND iSUN. “\*” INDICATES THAT THE RESULTS ARE OBTAINED VIA SELF-IMPLEMENTED NETWORKS. THE CURVATURE OF HiO IS SET TO  $-0.01$

Image encoder (on ID dataset)	Method	FPR (95% TPR) ↓	Detection Error ↓	AUROC ↑
Dense-BC (on CIFAR-10)	ODIN* [16]	11.2	6.8	98.0
	HiO	6.0	5.2	98.8
Dense-BC (on CIFAR-100)	ODIN* [16]	48.0	17.3	90.3
	HiO	40.9	16.5	91.1
WRN-28-10 (on CIFAR-10)	ODIN* [16]	22.9	10.1	95.8
	HiO	15.5	8.3	97.3
WRN-28-10 (on CIFAR-100)	ODIN* [16]	52.2	19.2	88.8
	HiO	49.5	20.0	88.3

TABLE V

MULTICLASS OOD DETECTION ON CIFAR-100 [61] WITH DENSENET [62] AND WRN-28-10 [63] IMAGE ENCODERS. MIXED-GEOMETRY EMBEDDING IN THE MiT MODEL COMBINES A SPHERICAL EMBEDDING AND HYPERBOLIC EMBEDDING. WE PROVIDE AVERAGING RESULTS OVER FOUR MULTIPLE OOD DATASETS: TINc, TINr, LSUNc, AND LSUNr. THE CURVATURES OF SiT, HiT, AND MiT ARE SET TO 1.0,  $-0.001$ , AND  $(1.0, -0.001)$ , RESPECTIVELY

Image encoder (on ID dataset)	Method	FPR (95% TPR) ↓	Detection Error ↓	AUROC ↑
Dense-BC (on CIFAR-100)	ELOC [35]	14.93	8.37	97.28
	SiT/HiT/MiT	17.84/10.45/12.84	9.20/6.78/7.75	96.72/98.05/97.63
WRN-28-10 (on CIFAR-100)	ELOC [35]	21.40	10.48	95.87
	SiT/HiT/MiT	16.91/13.63/17.73	8.75/7.87/9.10	97.01/97.47/96.78

## V. EXPERIMENTS

In this section, we evaluate our models on four visual anomaly, open-set, or OOD tasks: 1) multiclass OOD detection; 2) one-class anomaly detection; 3) multiclass anomaly segmentation; and 4) one-class anomaly segmentation. Table II

shows the difference between each task used in this article. For simplicity, we use the following abbreviations for our models: spherical-in-one (SiO), SiT, HiO, HiT, MiO, and MiT. It is notable that such terms (e.g., SiO and SiT) indicate that the curved geometric models apply the geometric classifiers

TABLE VI

ONE-CLASS ANOMALY DETECTION ON CIFAR-10 [61]. IMAGE-LEVEL AUROC IN % IS GIVEN. **class<sub>*i*</sub>** INDICATES THE *i*TH CLASS. “EXTRA” INDICATES UTILIZING EXTRA DATA FOR TRAINING (E.G., USING PRETRAINED MODELS ON IMAGENET [64]). “\*” INDICATES THAT THE RESULTS ARE OBTAINED VIA A SELF-IMPLEMENTED NETWORK. MIXED-GEOMETRY EMBEDDING IN THE MIT MODEL INCORPORATES BOTH THE SPHERICAL EMBEDDING AND HYPERBOLIC EMBEDDING. ROTNET\*, SiT, HiT, AND MiT ADOPT WRN-28-10 [63] AS IMAGE ENCODER, WHILE CSI, SiT, HiT, AND MiT UTILIZE RESNET-18 [65]. THE CURVATURES OF SiT, HiT, AND MiT BASED ON ROTNET\* ARE SET TO 1.0,  $-0.005$ , AND  $(1.0, -0.005)$ , RESPECTIVELY. THE CURVATURES OF SiT, HiT, AND MiT BASED ON CSI ARE GIVEN AS 1.0,  $-0.01$ , AND  $(1.0, -0.01)$ , RESPECTIVELY

Method	Extra	class <sub>0</sub>	class <sub>1</sub>	class <sub>2</sub>	class <sub>3</sub>	class <sub>4</sub>	class <sub>5</sub>	class <sub>6</sub>	class <sub>7</sub>	class <sub>8</sub>	class <sub>9</sub>	Avg
RotNet* [36]	✗	72.70	94.25	76.38	69.26	79.30	80.97	76.18	92.78	90.62	88.67	82.11
SiT	✗	73.77	95.21	78.39	70.46	80.92	81.75	77.80	93.25	91.27	89.40	83.22
HiT	✗	72.68	94.54	78.31	69.62	80.80	81.63	77.05	93.06	91.06	88.90	82.77
MiT	✗	73.60	95.18	79.82	70.80	81.76	82.57	78.52	93.52	91.16	89.46	<b>83.64</b>
CSI [38]	✗	89.9	99.1	93.1	86.4	93.9	93.2	95.1	98.7	97.9	95.5	94.3
SiT	✗	89.26	99.17	94.16	87.76	94.53	93.47	95.44	98.85	97.91	96.03	94.66
HiT	✗	89.40	99.17	94.22	88.06	94.33	93.48	95.58	98.76	97.85	95.93	94.68
MiT	✗	89.46	99.16	94.30	88.04	94.41	93.32	95.58	98.87	97.94	95.99	<b>94.71</b>

TABLE VII

MULTICLASS ANOMALY SEGMENTATION ON STREETHAZARDS [19]. FPR (95% TPR), PIXEL-LEVEL AUROC, AND AUPR IN % ARE GIVEN. THE METRIC STD OVER FIVE RUNS. THE RESULTS OF AE, DROPOUT, AND MSP ARE PROVIDED IN [19]. THE CURVATURES OF SiT, HiT, AND MiT BASED ON MSP ARE SET TO 1.0,  $-0.01$ , AND  $(1.0, -0.01)$ , RESPECTIVELY

Method	FPR (95% TPR) ↓	AUROC ↑	AUPR ↑
AE [66]	91.7	66.1	2.2
Dropout [67]	79.4	69.9	7.5
MSP [15]	33.7	87.7	6.6
SynthCP [25]	28.4	88.5	9.3
MaxLogit [19]	<b>26.5</b>	<b>89.3</b>	<b>10.6</b>
MSP [15]	33.7	87.7	6.6
SiT	29.9	86.4	4.6
HiT	27.7	88.9	6.3
MiT	<b>27.1</b>	<b>89.5</b>	<b>8.6</b>

compared to their corresponding baselines (see Table I for more details). Metrics, including the false positive rate at 95% true positive rate (FPR at 95% TPR), the detection error, the area under the receiver operating characteristics (AUROC) [72], [73], and the area under the precision-recall (AUPR) [74], [75], are measured. All results are averaged over five independent trials.

For the experiments on multiclass OOD detection, we train the geometric models for 300/300/100 epochs using Hendrycks&Gimpel [15]/ODIN [16]/ELOC [35] as the baselines, with a learning rate of 0.05/0.05/0.1. In the experiments of one-class anomaly detection, the models based on RotNet [36]/CSI [38] are trained for 100/1000 epochs with learning rates of 0.01/0.1. For the multiclass segmentation task, the models are trained for 30 epochs with a learning rate of 0.02. In one-class anomaly segmentation, the models are trained for 1200 epochs with a learning rate of  $1e^{-4}$ . The values of curvatures for the GiO/GiT models are provided in the captions of Tables III–IX.

### A. Multiclass OOD Detection

The objective of multiclass OOD detection, traditionally termed OOD detection, is to identify whether a sample is from the given dataset with multiple ID classes. The model

is trained on the ID dataset only. In this setting, CIFAR-10 and CIFAR-100 [61] are chosen as ID datasets, while the cropped TinyImageNet (TINc), the resized TinyImageNet (TINr) [64], the cropped LSUN (LSUNc), the resized LSUN (LSUNr) [76], and iSUN [77] are OOD datasets. We first adopt the Hendrycks&Gimpel model [15] as the baseline network. Both Dense-BC [62] and WRN-28-10 [63] are used as image encoders. As shown in Table III, we can observe that the HiO model attains the overall best accuracy. In addition, our models, except the SiO model, bring the performance gain over the baselines, showing the superiority of curved geometries as embedding spaces. It is also notable that our models are light. For example, the HiT model surpasses the baseline by 4.2% with WRN-28-10 on CIFAR-100, while it only uses an extra 0.02 M parameters, i.e., from 146.05 M to 146.07 M. Moreover, in most cases, the performance of the mixed-curvature model, MiO or MiT, is in between that of hyperbolic and spherical models. Besides the Hendrycks&Gimpel model, we also use ODIN [16] as the baseline where we employ the input preprocessing at the test phase. The results of geometric models that adopt ODIN as the baseline are reported in Table IV. From Table III, we identify HiO as the model, which obtains the best performance. Hence, we choose and test HiO for ODIN. Except for the experiment of WRN-28-10 on CIFAR-100, we see that HiO boosts the performance against the baseline.

In addition to Hendrycks&Gimpel and ODIN models, we also incorporate the proposed geometric classifier into advanced baselines. In this study, we employ the ELOC [35] as the baseline network. As shown in Table V, the HiT model performs the best over two image encoders. Specifically, it surpasses the baseline by 0.77%/1.60% in AUROC under Dense-BC/WRN-28-10. Similar to the results shown in Table III, the performance of the MiT model is in between SiT and HiT with Dense-BC on CIFAR-100, but it unexpectedly becomes the worst with WRN-28-10.

### B. One-Class Anomaly Detection

In the one-class anomaly (or open-set) detection setting, only one class is set as the normal class, while other classes

TABLE VIII

ONE-CLASS ANOMALY SEGMENTATION (ANOMALY LOCALIZATION) ON MVTECAD [68]. PIXEL-LEVEL AUROC IN % IS GIVEN. “EXTRA” INDICATES UTILIZING EXTRA DATA FOR TRAINING (E.G., USING PRETRAINED MODELS ON IMAGENET [64]). “CPR” INDICATES USING THE COMPARISON WITH THE TRAINING DATA FOR ANOMALY SCORE COMPUTATION DURING EVALUATION. THE RESULTS OF AVID AND ANOGAN/AE ARE PROVIDED IN [32] AND [49], RESPECTIVELY. MIXED-GEOMETRY EMBEDDING IN THE MI0/MI1 MODEL INCLUDES SPHERICAL EMBEDDING AND HYPERBOLIC EMBEDDING. THE CURVATURES OF SiO, HiO, MiO, SiT, HiT, AND MiT ARE SET TO 1.0,  $-0.01$ ,  $(1.0, -0.01)$ ,  $1.0, -1.0$ , AND  $(1.0, -1.0)$ , RESPECTIVELY. THE MODEL SIZES OF PATCH-SVDD<sup>†</sup>, SiO, HiO, MiO, SiT, AND MiT ARE 1.72 M, 1.72 M, 1.72 M, 1.82 M, 1.82 M, 1.83 M, AND 1.93 M, RESPECTIVELY

Method	Extra	Cpr	Bottle	Hazelnut	Capsule	Metal Nut	Leather	Pill	Wood	Carpet	Tile	Grid	Cable	Transistor	Toothbrush	Screw	Zipper	Avg
AnoGAN [69]	X	X	86	87	84	76	64	87	62	54	50	58	78	80	90	80	78	74
AVID [23]	X	X	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	77
AE-L2 [70]	X	X	86	95	88	86	75	85	73	59	51	90	86	86	93	96	77	82
AE-SSIM [70]	X	X	93	97	94	89	78	91	73	87	59	94	82	90	92	96	88	87
VAE-VE [33]	X	X	87	98	74	94	95	83	77	78	80	73	90	93	94	97	78	86
VAE-grad [71]	X	X	92.2	97.6	91.7	90.7	92.5	93	83.8	73.5	65.4	96.1	91.0	91.9	98.5	94.5	86.9	89.3
Patch-SVDD <sup>†</sup> [37]	X	X	62.78	71.03	80.03	68.28	90.03	67.15	65.45	82.60	89.28	82.55	67.60	50.35	88.20	85.95	67.23	74.57
SiO	X	X	62.80	73.53	80.50	68.40	89.70	70.13	65.50	74.10	85.90	82.10	68.35	51.10	86.57	50.00	68.35	71.80
HiO	X	X	68.27	76.43	89.00	50.00	93.65	73.20	79.90	86.40	86.65	78.50	87.45	71.50	87.03	78.65	90.60	79.82
MiO	X	X	68.50	79.73	92.03	65.28	94.80	74.83	79.00	86.20	70.20	66.17	61.93	64.47	87.43	71.10	88.08	76.65
SiT	X	X	95.43	94.60	90.80	96.07	96.30	92.20	82.93	89.20	92.55	89.13	93.67	92.40	93.30	83.40	91.68	91.58
HiT	X	X	95.80	93.60	92.87	88.40	97.03	92.90	81.10	83.13	87.03	92.03	95.20	95.27	94.57	81.63	94.20	90.98
MiT	X	X	96.33	96.05	92.77	96.77	96.63	93.50	82.30	84.30	93.63	93.70	94.93	94.40	95.03	77.33	94.13	<b>92.12</b>

are used as abnormal classes. The common practice of creating discriminative representations under this setting is modeled as a multiclass classification problem using the self-supervised learning (SSL) algorithms [36], [38], [38], [41], [42], [43]. In this task, we evaluate our models on the one-class CIFAR-10 dataset [61].

Our geometric classifier is built on RotNet [36] and CSI [38]. RotNet predicts the rotation angles as supervision for SSL. Following the setting in [36], we set the rotation angles to  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ . A 4-D classifier that predicts the angle of rotation is applied to the input image. CSI adopts the contrastive learning scheme, which contrasts the negative samples coming from the data augmentation. The results are shown in Table VI. We can observe that either of our models can improve the baselines, showing that embedding in curved spaces indeed benefits the discrimination of data embedding. For example, in RotNet, the method with mixed-curvature geometry, MiT, attains the best performance improvement, e.g., 1.53%, and outperforms the SiT and HiT models. It verifies that mixed-curvature geometry indeed benefits from the advantages of both spherical geometry and hyperbolic geometry. In CSI [38], as a strong baseline, our models again bring a performance gain, and the MiT method achieves the best average performance, revealing that our models generalize and are effective.

### C. Multiclass Anomaly Segmentation

In contrast to multiclass OOD detection, which recognizes the OOD samples at the image level, multiclass anomaly (or open-set) segmentation is required to predict anomalous objects at the pixel level. Following [19] and [25], we evaluate this task using the StreetHazards dataset [19]. The Hendrycks&Gimpel model [15] with maximum softmax probability (MSP) is adopted as the baseline. We report the results in Table VII. As suggested in Table VII, this task also benefits the most from the mixed-curvature geometry (MiT), again showing that multiple geometries are essential to learning discriminative embeddings.

### D. One-Class Anomaly Segmentation

One-class anomaly segmentation, also known as anomaly localization, aims to identify whether the input pixel is an anomalous pixel or not [33], [68]. In contrast to the multiclass anomaly segmentation, the training samples in one-class anomaly segmentation are drawn from only one class of the dataset.

We verify the effectiveness of our models on the MVTecAD dataset [68] and adopt the SOTA model Patch-SVDD [37] as a baseline without using extra data. A possible limitation of Patch-SVDD is that the computation of the anomaly scores AS for inference depends to a great extent on the comparison with training samples. To simplify the evaluation process, we calculate the anomaly score directly from its normalized classifier’s value without utilizing any training data (denoted by Patch-SVDD<sup>†</sup>). We then plug our geometric models on top of Patch-SVDD<sup>†</sup>. The results are reported in Table VIII. As suggested in Table VIII, all geometric models, except the SiO model, boost the performance of the baseline, and the mixed-curvature geometric model, MiT, performs the best. It gains 17.55% improvement. In this task, the dual-branch architecture (i.e., SiT/HiT/MiT) consistently outperforms the single-branch model (i.e., SiO/HiO/MiO). Along with a considerable improvement, our proposal is also cheap. For example, the SiT improves the baseline by a margin of 16.41%, while it only brings extra 0.1 M parameters, i.e., 1.82 M versus 1.72 M, again showing the benefits from curved geometric embeddings.

The idea of Patch-SVDD<sup>†</sup> is to evaluate our method on a toy example to illustrate the advantage of curved geometries. However, after we remove the comparison process with training images from Patch-SVDD, we find its identification performance significantly drops. To show the full potential of our design in conjunction with the original Patch-SVDD, we employ our geometric model over the original Patch-SVDD where training images are considered for calculating the anomaly score. As shown in Table IX, the accuracy of Patch-SVDD on MVTecAD is boosted from 95.7% to 96.5%/96.5%/96.7% once using SiT/HiT/MiT. We follow

TABLE IX

ONE-CLASS ANOMALY SEGMENTATION (ANOMALY LOCALIZATION) ON MVTECAD [68]. PIXEL-LEVEL AUROC IN % IS GIVEN. “EXTRA” INDICATES UTILIZING EXTRA DATA FOR TRAINING (E.G., USING PRETRAINED MODELS ON IMAGENET [64]). “CPR” INDICATES USING THE COMPARISON WITH THE TRAINING DATA FOR ANOMALY SCORE COMPUTATION DURING EVALUATION. THE RESULTS OF LSA ARE PROVIDED IN [32]. MIXED-GEOMETRY EMBEDDING IN THE MiT MODEL INCLUDES SPHERICAL EMBEDDING AND HYPERBOLIC EMBEDDING. THE CURVATURES OF SiT, HiT, AND MiT ARE SET TO 1.0,  $-1.0$ , AND  $(1.0, -1.0)$ , RESPECTIVELY

Method	Extra	Cpr	Bottle	Hazelnut	Capsule	Metal Nut	Leather	Pill	Wood	Carpet	Tile	Grid	Cable	Transistor	Toothbrush	Screw	Zipper	Avg
LSA [28]	✓	✗	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	81
CAVGA-Rw [32]	✓	✗	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	90
MR-KD [49]	✓	✗	96.32	94.62	95.86	86.38	98.05	89.63	84.80	95.64	82.77	91.78	82.40	76.45	96.12	95.96	93.90	90.71
Glancing-at-Patch [50]	✓	✗	93	84	90	91	90	93	81	96	80	78	94	100	96	96	99	91
PANDA-OE [78]	✓	✗	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	96.2
CutPaste [45]	✗	✓	97.6	97.3	97.4	93.1	99.5	95.7	95.5	98.3	90.5	97.5	90.0	93.0	98.1	96.7	99.3	96.0
Patch-SVDD [37]	✗	✓	98.1	97.5	95.8	98.0	97.4	95.1	90.8	92.6	91.4	96.2	96.8	97.0	98.1	95.7	95.1	95.7
SiT	✗	✓	98.10	98.93	97.00	98.00	97.20	94.83	92.30	91.20	96.30	97.43	96.30	97.83	96.40	97.53	98.45	96.52
HiT	✗	✓	98.23	96.90	96.83	97.60	97.20	94.20	92.73	94.13	96.50	97.65	95.75	97.60	95.95	97.37	98.37	96.47
MiT	✗	✓	98.23	99.00	96.33	98.80	97.20	96.03	92.43	91.70	96.60	97.50	96.80	97.63	96.65	97.73	98.45	96.74

TABLE X

MULTICLASS ANOMALY SEGMENTATION ON MVTECAD [68]. PIXEL-LEVEL AUROC IN % IS GIVEN. MIXED-GEOMETRY EMBEDDING IN THE MiT MODEL INCLUDES SPHERICAL EMBEDDING AND HYPERBOLIC EMBEDDING. THE CURVATURES OF SiT, HiT, AND MiT ARE SET TO 1.0,  $-1.0$ , AND  $(1.0, -1.0)$ , RESPECTIVELY

Method	Bottle	Hazelnut	Capsule	Metal Nut	Leather	Pill	Wood	Carpet	Tile	Grid	Cable	Transistor	Toothbrush	Screw	Zipper	Avg
Patch-SVDD <sup>†</sup> [37]	67.8	77.1	89.1	79.1	62.8	75.9	93.6	61.9	80.2	84.3	84.7	82.6	74.3	88.2	72.2	78.3
SiT	84.5	92.6	89.8	86.8	97.9	83.0	88.7	88.0	88.7	70.2	77.9	72.3	81.6	77.1	91.3	84.7
HiT	85.1	92.9	88.9	85.0	97.2	85.2	86.6	91.3	91.6	76.6	74.1	84.1	81.2	63.8	91.1	85.0
MiT	83.5	90.7	89.2	88.5	98.0	87.3	86.1	84.6	92.6	70.8	76.3	74.6	82.7	63.8	91.3	84.0

the anomaly score computation of the original Patch-SVDD except for replacing the patch’s embedding with the geometric score  $z_{EG}$ .

As described in Section V-C, the multiclass setting allows training data from multiple classes. We utilize this multiclass data to train PatchSVDD<sup>†</sup> and the proposed geometric models on the MVTECAD dataset [68]. Consistent with the approach outlined in [79], we do not employ class label information. The results, presented in Table X, demonstrate the clear superiority of the geometric approaches over the baseline model.

### E. Analysis

In this section, we aim to provide studies to analyze the superiority of our design.

1) *Performance*: We learn from empirical observations in Sections V-A–V-D that the curved spaces are able to consistently provide reliable information for anomaly, open-set, or OOD recognition. In most cases, our curvature-aware geometric networks clearly outperform Euclidean networks. One possible explanation is that the geometric representations benefit particular problems. For example, as we discussed in Section I, hyperbolic geometry is good at encoding hierarchical structures inside the data. The datasets we test include hierarchical information to some extent. For instance, the CIFAR dataset includes ten superclasses and ten subclasses under each superclass. The 15 classes of MVTECAD can be categorized into two main supercategories, “object” and “texture.” Hence, the hyperbolic space takes effect.

2) *Mixed Geometry*: Another interesting fact has been observed is that the mixed-curvature geometry beats its component single-curvature geometries in several cases: MiT based on RotNet\*/CSI of one-class anomaly detection, MiT based on MSP of multiclass anomaly segmentation, and MiT based on Patch-SVDD<sup>†</sup> of one-class anomaly segmentation in Tables VI–VIII, respectively. In some cases,

TABLE XI

ABLATION STUDY: MULTICLASS OOD DETECTION. IMAGE-LEVEL AUROC IN % IS GIVEN. DIFFERENT CASES ARE EVALUATED ON CIFAR-10 [61] TO VERIFY THE INTERACTIONS BETWEEN HYPERBOLIC OR SPHERICAL AND EUCLIDEAN SPACES

HiO	HiT	$e_E$ in HiT	$e_H$ in HiT
97.1	94.1	92.1	94.5
SiO	SiT	$e_E$ in SiT	$e_S$ in SiT
93.4	97.1	93.6	96.8

the mixed-curvature geometry has a balanced performance. For example, in the task of multiclass OOD detection, there exists a significant performance gap between hyperbolic and spherical geometries, as evidenced by SiO versus HiO. Thus, the mixed space MiO might be expected to have an average performance.

3) *Interactions Among Geometries*: The GiT model requires meanwhile learning two embeddings (e.g., a Euclidean  $e_E$ , and a hyperbolic  $e_H$  or spherical embedding  $e_S$ ). We observe that it happens in the interactions between different geometric components. For instance, in multiclass OOD detection (see Table III),  $e_E$  could enrich  $e_S$  (SiO versus SiT); however, for  $e_H$ , it has less or even negative impact (HiO versus HiT). To further understand the influence, we analyzed the experiments of WRN-28-10 on CIFAR-10 where we separately test  $e_E$  and  $e_H$  (or  $e_S$ ) in HiT (or SiT). Results in Table XI suggest the aforementioned point (HiO versus  $e_H$  in HiT and SiO versus  $e_S$  in SiT).

4) *Curvature*: The curvature  $\kappa$  is the only hyperparameter in the proposed curvature-aware geometric networks. The study of one-class anomaly segmentation of HiT based on Patch-SVDD<sup>†</sup> in Table XII suggests that  $\kappa$  clearly has an impact on anomaly recognition performance. Table XII shows that for each category, there exists an optimal curvature value  $\kappa$  that leads to the best performance. Deviating from this optimal value results in a decrease in performance.

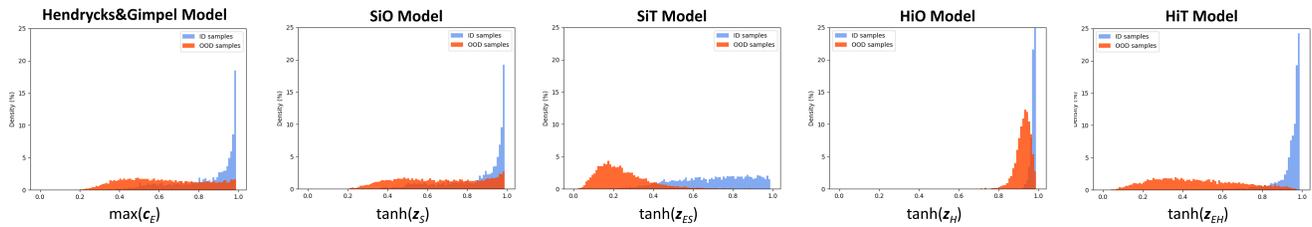


Fig. 5. Visualization: multiclass OOD detection. Density distributions of  $\max(\mathbf{c}_E)$ ,  $\tanh(z_S)$ ,  $\tanh(z_{ES})$ ,  $\tanh(z_H)$ , and  $\tanh(z_{EH})$  of Hendrycks&Gimpel, SiO, SiT, HiO, and HiT models with Dense-BC on CIFAR10→TINc are provided. The obtained AUROC for these five models are 94.8%, 94.7%, 98.1%, 98.7%, and 96.0%. More corresponding results can be referred to in Table III.

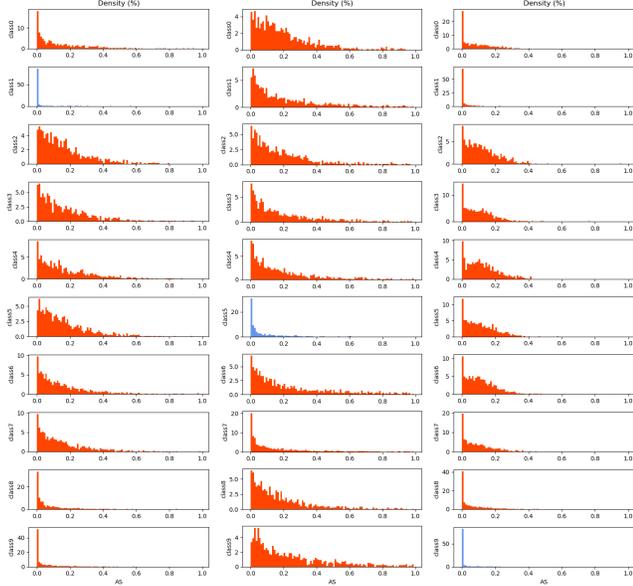


Fig. 6. Visualization: one-class anomaly detection. Density distribution of anomaly score  $AS(z_{EH})$  of HiT based on RotNet\* on CIFAR-10 [61] is visualized. We choose cases where **class1**, **class5**, or **class9** is taken as the normal class. The distribution of one case is shown in one column. More corresponding results can be referred to in Table VI.

5) *Method Choice*: Our comprehensive empirical study suggests that a single definitive conclusion cannot be made. This is in line with observations made in recent works. For example, in [14], the best geometry choice depends on the task. Our study clearly shows that curved geometry is beneficial in capturing the geometry of data, contributing tangibly to identifying anomalies in data. Specifically, our empirical study suggests that the preferred model for each task is shown as follows—multiclass OOD detection: HiO/HiT; multiclass anomaly segmentation: MiT; one-class anomaly detection: MiT; and one-class anomaly segmentation: MiT. If one model should be chosen in all instances, then we will opt for MiT as the potential model for the visual anomaly recognition tasks.

## F. Visualization

In this section, we qualitatively study our method in image-level classification task and pixel-level segmentation task, to understand why our method can bring performance gain over the baseline model.

1) *Image-Level Classification*: In this work, we particularly study whether the geometric score  $z_G$  and  $z_{EG}$  from the curved

TABLE XII  
ABLATION STUDY: ONE-CLASS ANOMALY SEGMENTATION (ANOMALY LOCALIZATION). PIXEL-LEVEL AUROC IN % IS GIVEN. DIFFERENT FIXED CURVATURES ON MVTECAD [68] ARE EVALUATED. FOR EACH CATEGORY, DIFFERENT CURVATURES BRING ABOUT DIFFERENT PERFORMANCES. ALSO, THE OPTIMAL CURVATURE CHOICE IS NOT CONSISTENT AMONG DIFFERENT CATEGORIES

Method	$\kappa$	Hazelnut	Leather	Wood	Toothbrush
HiT	-1e-5	85.53	96.31	79.24	88.26
HiT	-1e-4	93.77	96.30	<b>84.67</b>	89.90
HiT	-1e-2	<b>95.40</b>	95.77	83.87	88.83
HiT	-1.0	93.60	<b>97.03</b>	81.10	<b>94.57</b>
HiT	-2.0	84.52	96.08	81.79	88.51

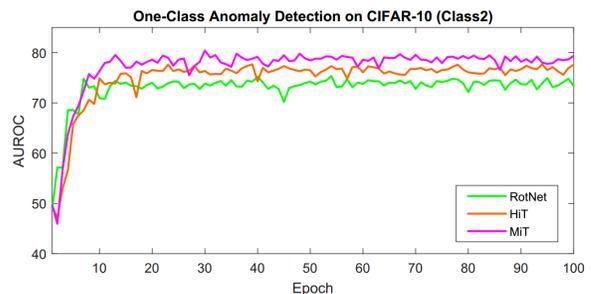
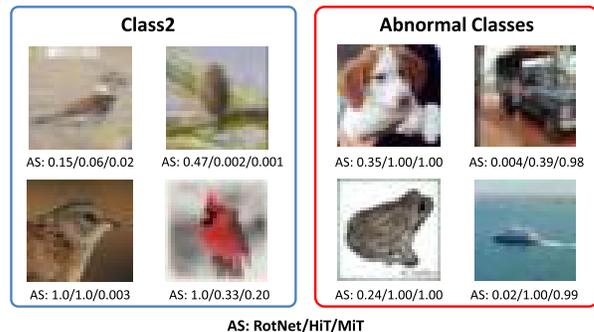


Fig. 7. Visualization: one-class anomaly detection. Examples with anomaly scores AS computed by RotNet\*, HiT, and MiT are provided (see the top). In addition, AUROC curves along the training epoch are plotted in the bottom.

embedding spaces can provide more useful information than the confidence vector  $\mathbf{c}_E$  from the Euclidean space in distinguishing normal (or closed-set/ID) and abnormal (or open-set/OOD) objects. We visualize the distribution of  $\max(\mathbf{c}_E)$ ,  $\tanh(z_S)$ ,  $\tanh(z_{ES})$ ,  $\tanh(z_H)$ , and  $\tanh(z_{EH})$  on multiclass

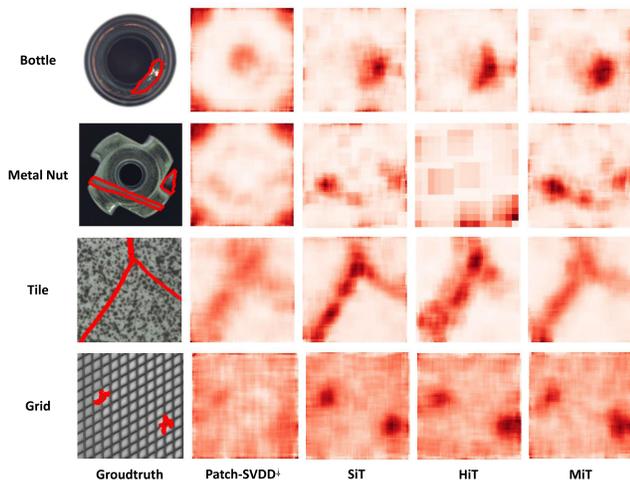


Fig. 8. Visualization: one-class anomaly segmentation (anomaly localization). The heatmaps of anomaly score  $AS(\mathbf{c}_E)$ ,  $AS(z_{ES})$ ,  $AS(z_{EH})$ , and  $AS(z_{EM})$  of Patch-SVDD<sup>†</sup>, SiT, HiT, and MiT on MVTecAD [68] are visualized. We provide examples from “bottle,” “metal nut,” “tile,” and “grid.” More corresponding results can be found in Table VIII. In addition, for the category “bottle,” AUROC curves of different models along the training epoch are plotted in the bottom.

OOD detection over CIFAR-10 (CIFAR10→TINc) in Fig. 5. As shown in Fig. 5, the curved embedding spaces help to better separate ID and OOD distributions. The distribution of anomaly score  $AS(z_{EH})$  of one-class anomaly detection on one-class CIFAR-10 is visualized in Fig. 6. The visualization shows that  $z_{EH}$  provides reliable information for distinguishing the normal one-class and the abnormal classes. We present some examples in Fig. 7 to verify the performance differences among models, where we compare RotNet\*/HiT/MiT on Class 2 as a normal class. The anomaly scores AS in the top figure show different models that have different capacities to recognize abnormal classes. Also, the AUROC curve suggests substantial improvements led by the curved geometries.

2) *Pixel-Level Segmentation*: From experiments, we find that besides image-level tasks, curved spaces as embedding spaces do help in pixel-level anomaly tasks. In Fig. 8, we show examples in which we visualize the anomaly score AS of Patch-SVDD<sup>†</sup>, SiT, HiT, and MiT on MVTecAD [68]. As shown in Fig. 8,  $z_{EG}$  from  $\mathcal{M}_G$  outperforms  $\mathbf{c}_E$  from  $\mathcal{M}_E$  in identifying anomalous pixels.

## VI. CONCLUSION

In this article, we study the potential use and benefit of employing curved spaces for the purpose of visual anomaly, open-set, or OOD recognition tasks. Our idea is inspired by the observation that curved embedding spaces help better represent “unknown” data in low-shot problems. Our work proposes two novel geometric networks, GiO and GiT, for the visual anomaly data analysis. In each geometric model, we fully study the potential of different geometry constraints. To the best of our knowledge, our curvature-aware geometric networks are the first attempt to employ curved geometries in visual anomaly, open-set, or OOD recognition. Based on extensive experiments, we confirm that more distinct representations between normal (or closed-set/ID) and anomalous (or open-set/OOD) samples can be learned using curved spaces, clearly showing the benefits of the curved spaces. We hope that this work can inspire researchers to explore curved geometries further in other domains.

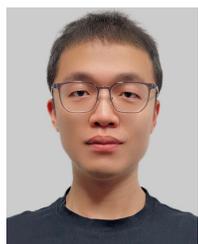
While the proposed geometric modules successfully enhance performance, their applicability is currently limited to baselines employing the “encoder–classifier” structure. Furthermore, the fixed curvature of these designs does not guarantee the optimal performance. To address these limitations, future research could explore the integration of curved embeddings into generative-based models with “encoder–decoder” structures. In addition, efforts can be directed toward developing adaptive-curvature designs to achieve the optimal performance.

## REFERENCES

- [1] M. Nickel and D. Kiela, “Poincaré embeddings for learning hierarchical representations,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 6338–6347.
- [2] A. Tifrea, G. Bécigneul, and O.-E. Ganea, “Poincaré GloVe: Hyperbolic word embeddings,” 2018, *arXiv:1810.06546*.
- [3] B. Dhingra, C. J. Shallue, M. Norouzi, A. M. Dai, and G. E. Dahl, “Embedding text in hyperbolic spaces,” 2018, *arXiv:1806.04313*.
- [4] V. Khruikov, L. Mirvakhabova, E. Ustinova, I. Oseledets, and V. Lempitsky, “Hyperbolic image embeddings,” in *Proc. IEEE/CVF Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6418–6428.
- [5] J. Park, J. Cho, H. J. Chang, and J. Young Choi, “Unsupervised hyperbolic representation learning via message passing auto-encoders,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 5516–5526.
- [6] Y. Zhang, L. Luo, W. Xian, and H. Huang, “Learning better visual data similarities via new grouplet non-Euclidean embedding,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 9918–9927.
- [7] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, “SphereFace: Deep hypersphere embedding for face recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 212–220.
- [8] X. Fan, W. Jiang, H. Luo, and M. Fei, “SphereReID: Deep hypersphere manifold embedding for person re-identification,” *J. Vis. Commun. Image Represent.*, vol. 60, pp. 51–58, Apr. 2019.
- [9] S. Ravi and H. Larochelle, “Optimization as a model for few-shot learning,” in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–11.
- [10] L. Van der Maaten and G. Hinton, “Visualizing data using t-SNE,” *J. Mach. Learn. Res.*, vol. 9, no. 11, pp. 2579–2605, 2008.
- [11] S. Liu, J. Chen, L. Pan, C.-W. Ngo, T.-S. Chua, and Y.-G. Jiang, “Hyperbolic visual embedding learning for zero-shot recognition,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 9273–9281.
- [12] J. Shen, Z. Xiao, X. Zhen, and L. Zhang, “Spherical zero-shot learning,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 2, pp. 634–645, Feb. 2022.

- [13] P. Fang, M. Harandi, and L. Petersson, "Kernel methods in hyperbolic spaces," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 10665–10674.
- [14] O. Skopek, O.-E. Ganea, and G. Bécigneul, "Mixed-curvature variational autoencoders," in *Proc. ICLR*, 2020, pp. 1–44.
- [15] D. Hendrycks and K. Gimpel, "A baseline for detecting misclassified and out-of-distribution examples in neural networks," in *Proc. ICLR*, 2017, pp. 1–12.
- [16] S. Liang, Y. Li, and R. Srikant, "Enhancing the reliability of out-of-distribution image detection in neural networks," in *Proc. ICLR*, 2018, pp. 1–27.
- [17] T. DeVries and G. W. Taylor, "Learning confidence for out-of-distribution detection in neural networks," 2018, *arXiv:1802.04865*.
- [18] C. Corbière, N. Thome, A. Bar-Hen, M. Cord, and P. Pérez, "Addressing failure prediction by learning model confidence," in *Proc. NeurIPS*, 2019, pp. 1–12.
- [19] D. Hendrycks et al., "Scaling out-of-distribution detection for real-world settings," in *Proc. 39th Int. Conf. Mach. Learn. (ICML)*, 2022, pp. 8759–8773.
- [20] Y.-C. Hsu, Y. Shen, H. Jin, and Z. Kira, "Generalized ODIN: Detecting out-of-distribution image without learning from out-of-distribution data," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 10951–10960.
- [21] J. Jang and C. O. Kim, "Collective decision of one-vs-rest networks for open-set recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Jul. 14, 2022, doi: 10.1109/TNNLS.2022.3189996.
- [22] K. Lee, H. Lee, K. Lee, and J. Shin, "Training confidence-calibrated classifiers for detecting out-of-distribution samples," in *Proc. ICLR*, 2018, pp. 1–16.
- [23] M. Sabokrou et al., "AVID: Adversarial visual irregularity detection," in *Proc. Asian Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 488–505.
- [24] K. Lis, K. K. Nakka, P. Fua, and M. Salzmann, "Detecting the unexpected via image resynthesis," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2152–2161.
- [25] Y. Xia, Y. Zhang, F. Liu, W. Shen, and A. L. Yuille, "Synthesize then compare: Detecting failures and anomalies for semantic segmentation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 145–161.
- [26] S. Kong and D. Ramanan, "OpenGAN: Open-set recognition via open data generation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 813–822.
- [27] J. Ren et al., "Likelihood ratios for out-of-distribution detection," in *Proc. NeurIPS*, 2019, pp. 1–12.
- [28] D. Abati, A. Porrello, S. Calderara, and R. Cucchiara, "Latent space autoregression for novelty detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 481–490.
- [29] P. Perera, R. Nallapati, and B. Xiang, "OCGAN: One-class novelty detection using GANs with constrained latent representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2898–2906.
- [30] D. Gong et al., "Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1705–1714.
- [31] F. V. Massoli, F. Falchi, A. Kantarci, S. Akti, H. K. Ekenel, and G. Amato, "MOCCA: Multilayer one-class classification for anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2313–2323, Jun. 2022.
- [32] S. Venkataramanan, K.-C. Peng, R. V. Singh, and A. Mahalanobis, "Attention guided anomaly localization in images," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 485–503.
- [33] W. Liu et al., "Towards visually explaining variational autoencoders," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8642–8651.
- [34] K. Zhou et al., "Memorizing structure-texture correspondence for image anomaly detection," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2335–2349, Jun. 2022.
- [35] A. Vyas, N. Jammalamadaka, X. Zhu, D. Das, B. Kaul, and T. L. Willke, "Out-of-distribution detection using an ensemble of self supervised leave-out classifiers," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 550–564.
- [36] S. Gidaris, P. Singh, and N. Komodakis, "Unsupervised representation learning by predicting image rotations," in *Proc. ICLR*, 2018, pp. 1–14.
- [37] J. Yi and S. Yoon, "Patch SVDD: Patch-level SVDD for anomaly detection and segmentation," in *Proc. Asian Conf. Comput. Vis.*, 2020, pp. 1–16.
- [38] J. Tack, S. Mo, J. Jeong, and J. Shin, "CSI: Novelty detection via contrastive learning on distributionally shifted instances," in *Proc. NeurIPS*, 2020, pp. 1–14.
- [39] J. Zheng, W. Li, J. Hong, L. Petersson, and N. Barnes, "Towards open-set object detection and discovery," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2022, pp. 3960–3969.
- [40] J. Hong et al., "GOSS: Towards generalized open-set semantic segmentation," 2022, *arXiv:2203.12116*.
- [41] L. Ruff et al., "Deep one-class classification," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 4393–4402.
- [42] I. Golan and R. El-Yaniv, "Deep anomaly detection using geometric transformations," in *Proc. NeurIPS*, 2018, pp. 1–12.
- [43] D. Hendrycks, M. Mazeika, S. Kadavath, and D. Song, "Using self-supervised learning can improve model robustness and uncertainty," in *Proc. NeurIPS*, 2019, pp. 1–13.
- [44] L. Bergman and Y. Hoshen, "Classification-based anomaly detection for general data," in *Proc. ICLR*, 2020, pp. 1–12.
- [45] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "CutPaste: Self-supervised learning for anomaly detection and localization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 9659–9669.
- [46] Y. Liu, Z. Li, S. Pan, C. Gong, C. Zhou, and G. Karypis, "Anomaly detection on attributed networks via contrastive self-supervised learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 6, pp. 2378–2392, Jun. 2022.
- [47] Q. Yu and K. Aizawa, "Unsupervised out-of-distribution detection by maximum classifier discrepancy," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 9517–9525.
- [48] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 4182–4191.
- [49] M. Salehi, N. Sadjadi, S. Baselizadeh, M. H. Rohban, and H. R. Rabiee, "Multiresolution knowledge distillation for anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 14897–14907.
- [50] S. Wang, L. Wu, L. Cui, and Y. Shen, "Glancing at the patch: Anomaly localization with global and local feature comparison," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 254–263.
- [51] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 18–42, Jul. 2017.
- [52] M. Zaheer et al., "Deep sets," in *Proc. NeurIPS*, 2017, pp. 1–11.
- [53] P. Fang, P. Ji, L. Petersson, and M. Harandi, "Set augmented triplet loss for video person re-identification," in *Proc. WACV*, Jan. 2021, pp. 464–473.
- [54] A. Cheraghian et al., "Synthesized feature based few-shot class-incremental learning on a mixture of subspaces," in *Proc. ICCV*, Oct. 2021, pp. 8661–8670.
- [55] C. Simon, P. Koniusz, R. Nock, and M. Harandi, "Adaptive subspaces for few-shot learning," in *Proc. CVPR*, Jun. 2020, pp. 1–10.
- [56] A. M. Bronstein, M. M. Bronstein, and R. Kimmel, "Expression-invariant face recognition via spherical embedding," in *Proc. IEEE Int. Conf. Image Process.*, vol. 3, Sep. 2005, p. 756.
- [57] W. Liu et al., "Deep hyperspherical learning," 2017, *arXiv:1711.03189*.
- [58] R. Ma, P. Fang, T. Drummond, and M. Harandi, "Adaptive poincaré point to set distance for few-shot classification," in *Proc. AAAI Conf. Artif. Intell.*, vol. 36, no. 2, 2022, pp. 1926–1934.
- [59] S. Jayasumana, S. Ramalingam, and S. Kumar, "Model-efficient deep learning with kernelized classification," in *Proc. ICLR*, 2022, pp. 1–18. [Online]. Available: <https://openreview.net/forum?id=30SXt3-vvnM>
- [60] A. Gu, F. Sala, B. Guelin, and C. Ré, "Learning mixed-curvature representations in product spaces," in *Proc. ICLR*, 2019, pp. 1–21.
- [61] A. Krizhevsky et al., "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep., 2009.
- [62] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [63] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016, *arXiv:1605.07146*.

- [64] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [65] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [66] C. Baur, B. Wiestler, S. Albarqouni, and N. Navab, "Deep autoencoding models for unsupervised anomaly segmentation in brain MR images," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, pp. 161–169.
- [67] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.
- [68] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9584–9592.
- [69] T. Schlegl et al., "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Proc. Int. Conf. Inf. Process. Med. Imag. Cham, Switzerland: Springer*, 2017, pp. 146–157.
- [70] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," 2018, *arXiv:1807.02011*.
- [71] D. Dehaene, O. Frigo, S. Combexelle, and P. Eline, "Iterative energy-based projection on a normal data manifold for anomaly localization," in *Proc. ICLR*, 2020, pp. 1–17.
- [72] J. Davis and M. Goadrich, "The relationship between precision-recall and ROC curves," in *Proc. 23rd Int. Conf. Mach. Learn. (ICML)*, 2006, pp. 233–240.
- [73] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006.
- [74] C. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. Cambridge, MA, USA: MIT Press, 1999.
- [75] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0118432.
- [76] F. Yu, A. Seff, Y. Zhang, S. Song, T. Funkhouser, and J. Xiao, "LSUN: Construction of a large-scale image dataset using deep learning with humans in the loop," 2015, *arXiv:1506.03365*.
- [77] P. Xu, K. A. Ehinger, Y. Zhang, A. Finkelstein, S. R. Kulkarni, and J. Xiao, "TurkerGaze: Crowdsourcing saliency with webcam based eye tracking," 2015, *arXiv:1504.06755*.
- [78] T. Reiss, N. Cohen, L. Bergman, and Y. Hoshen, "PANDA: Adapting pretrained features for anomaly detection and segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2805–2813.
- [79] Z. You et al., "A unified model for multi-class anomaly detection," in *Proc. Adv. Neural Inf. Process. Syst.*, 2022, pp. 1–14.



**Jie Hong** received the M.Eng. degree from the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore, in 2017. He is currently pursuing the Ph.D. degree with the College of Engineering, Computing and Cybernetics, Australian National University (ANU), Canberra, ACT, Australia, and DATA61-CSIRO, Canberra. His research interests include computer vision, deep learning, robotics, and control systems.



**Pengfei Fang** received the M.E. degree from Australian National University (ANU), Canberra, ACT, Australia, in 2017, and the Ph.D. degree from ANU and DATA61-CSIRO, Canberra, in 2022.

He was a Post-Doctoral Fellow at Monash University, Clayton, VIC, Australia, in 2022. He is currently an Associate Professor with the School of Computer Science and Engineering, Southeast University (SEU), Nanjing, China.

His research interests include computer vision and machine learning.



**Weihao Li** received the Ph.D. degree in computer science from Heidelberg University, Heidelberg, Germany, in 2019, under the supervision of Prof. Dr. Carsten Rother.

He is currently a Research Fellow at Australian National University (ANU), Canberra, ACT, Australia. Before joining ANU in 2022, he was a Post-Doctoral Research Fellow at Data61-CSIRO, Canberra, from 2019 to 2022. His research interests include computer vision and machine learning.



**Junlin Han** received the Bachelor of Information Technology degree (Hons.) from Australian National University (ANU), Canberra, ACT, Australia, in 2023.

His academic pursuits focus on computer vision, deep learning, and artificial intelligence, with a specific emphasis on leveraging data-centric methodologies to attain human-like visual intelligence.

Mr. Han actively participates as a reviewer for esteemed conferences. He was honored as a top reviewer in Conference on Neural Information Processing Systems (NeurIPS) 2022.



**Lars Petersson** is currently the Group Leader and a Principal Research Scientist with the Imaging and Computer Vision Group, Data61-CSIRO, Canberra, ACT, Australia. He is also leading one of the activities under CSIRO's Machine Learning and Artificial Intelligence Future Science Platform effort where data science problems from the smallest of microscopy scales to the largest of astronomical scales are addressed. Previous to joining Data61-CSIRO, he was a Principal Researcher and the Research Leader of the Computer Vision Research

Group, NICTA, Canberra, where he was leading projects, such as smart cars, AutoMap, and distributed large-scale vision.



**Mehrtash Harandi** is currently an Associate Professor with the Department of Electrical and Computer Systems Engineering, Monash University, Clayton, VIC, Australia. He is also a contributing Research Scientist with the Machine Learning Research Group (MLRG), Data61-CSIRO, Canberra, ACT, Australia, and an Associate Investigator with the Australian Center for Robotic Vision (ACRV), Brisbane, QLD, Australia. His current research interests include theoretical and computational methods in machine learning, computer vision, signal processing, and Riemannian geometry.