

Reflection and Rotation Symmetry Detection via Equivariant Learning

Ahyun Seo Byungjin Kim Suha Kwak Minsu Cho

Pohang University of Science and Technology (POSTECH), South Korea

<http://cvlab.postech.ac.kr/research/EquiSym>

Abstract

The inherent challenge of detecting symmetries stems from arbitrary orientations of symmetry patterns; a reflection symmetry mirrors itself against an axis with a specific orientation while a rotation symmetry matches its rotated copy with a specific orientation. Discovering such symmetry patterns from an image thus benefits from an equivariant feature representation, which varies consistently with reflection and rotation of the image. In this work, we introduce a group-equivariant convolutional network for symmetry detection, dubbed EquiSym, which leverages equivariant feature maps with respect to a dihedral group of reflection and rotation. The proposed network is built end-to-end with dihedrally-equivariant layers and trained to output a spatial map for reflection axes or rotation centers. We also present a new dataset, *DENSE* and *DIVERSE* symmetry (*DENDI*), which mitigates limitations of existing benchmarks for reflection and rotation symmetry detection. Experiments show that our method achieves the state of the arts in symmetry detection on *LDRS* and *DENDI* datasets.

1. Introduction

From molecules to galaxies, from nature to man-made environments, symmetry is everywhere. Comprehensive perception and exploitation of real-world symmetry are the instinctive abilities of humans and animals that have the potential to take intelligent systems to the next level. The focus of this paper is on the two most primitive symmetries, reflection and rotation symmetries. The goal of reflection and rotation symmetry detection is to find a reflection axis and a rotation center that remain invariant under reflection and rotation, respectively. Despite decades of efforts [29, 46], symmetry detection methods have been limited to the well-defined symmetry patterns, and the remedy for real-world symmetry is still yet to be thoroughly explored. The simplicity of mathematical concepts of symmetry encouraged early approaches to find keypoint pairs that satisfy pre-defined constraints for symmetry [1, 3, 32, 37, 39, 43],

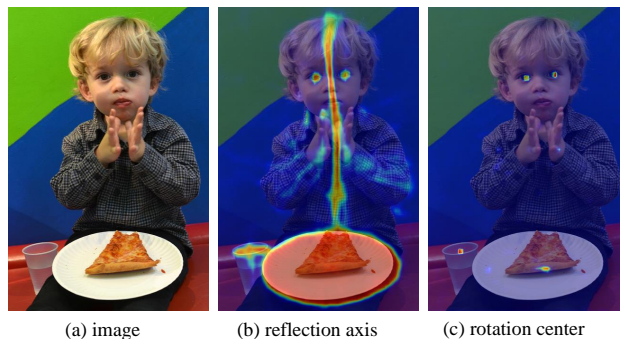


Figure 1. Symmetry detection examples of our method EquiSym. (a) an input image, (b) a score map of reflection symmetry axes, and (c) that of rotation symmetry centers. Best viewed in color.

which leverage hand-crafted local feature descriptors to detect sparse symmetry patterns. Recently, convolutional neural networks (CNNs) have been successfully applied to detect reflection symmetry and have surpassed the previous methods by learning score map regression [13] or symmetric matching [38] from data.

The primary challenge in detecting symmetry patterns lies in the fact that a symmetry manifests itself with an arbitrary orientation and perceiving the pattern requires an analysis based on the orientation; a reflection symmetry mirrors itself against an axis with a specific orientation and a rotation symmetry matches its rotated copy with a specific orientation. Most methods for symmetry detection thus involve searching over the space of candidate orientations of symmetry patterns and also developing a robust representation that is either *invariant* or *equivariant* with respect to rotation and reflection. The early approaches leverage an equivariant representation by extracting oriented keypoints and performing orientation normalization [1, 32, 37, 39, 43]. While this technique has proven effective for shallow gradient-based features, it cannot be applied to deep feature maps from standard neural networks, where rotation and reflection induce unpredictable variations in representation.

To address the challenge, we propose to learn a group-

equivariant convolutional neural network for reflection and rotation symmetry detection, dubbed *EquiSym*. Recently, there has been active research on equivariant networks to incorporate equivariance explicitly for robust and sample-efficient representation learning [7, 9, 19, 40, 44, 47]. Unlike standard neural networks, they induce predictable and structure-preserving representation with respect to the geometric transformations, *e.g.*, rotation or reflection, which is eminently suitable for symmetry detection. To detect consistent symmetry patterns over different orientations, we build a dihedrally-equivariant convolutional network [44], which is designed to be end-to-end equivariant to a group of reflection and rotation. The network effectively learns to output a score map of reflection axes for reflection symmetry or that of rotation centers for rotation symmetry.

We also present a new dataset, *DENse and DIverse symmetry* (DENDI), for reflection and rotation symmetry detection. DENDI contains real-world images with accurate and clean annotations for reflection and rotation symmetries and mitigates limitations of existing benchmarks [4, 12, 13, 27, 38]. First, the reflection symmetry axes are diverse in scale and orientation, while previous datasets mostly focus on the dominant axes of the vertical or horizontal ones. Second, the rotation centers are annotated to the objects in polygon and ellipse shape, not limited to the circular objects. Third, the number of the rotation folds for each rotation center is annotated, which is the first in a large-scale dataset. Finally, the number of images is 1.7x and 2.0x larger than the second-largest reflection and rotation symmetry detection datasets, respectively.

The contribution of our work can be summarized as:

- We propose a novel group-equivariant symmetry detection network, *EquiSym*, which outputs group-equivariant score maps for reflection axes or rotation centers via end-to-end reflection- and rotation-equivariant feature maps.
- We present a new dataset, *DENse and DIverse symmetry* dataset (DENDI), containing images of reflection and rotation symmetries annotated in a broader range of typical real-world objects.
- We show the outstanding performance of *EquiSym* in reflection symmetry detection on SDRW [27], LDRS [38], and DENDI, and in rotation symmetry detection on DENDI.

2. Related Work

2.1. Equivariant deep learning

Equivariance is a desirable inductive bias that improves generalization and sampling efficiency. The conventional convolution is equivariant to translations but not to other transformations such as rotations and reflections. Group

equivariant CNNs [7, 19] use group convolution to learn equivariant representations for symmetry groups. Marcos *et al.* [33] generate and propagate vector fields that maintain the maximum response along with the corresponding direction throughout the network. Worrall *et al.* [47] exploit circular harmonics to obtain rotational equivariance in a continuous domain. Cohen *et al.* [9] combine fixed base filters linearly, resulting in steerable filters with no interpolation artifacts. Equivariant CNNs on homogeneous spaces [5, 6, 8, 45] are also proposed. The aforementioned methods consider equivariance to specific transformations until Weiler *et al.* [44] provide a general solution of kernel space constraint for arbitrary group representations of the Euclidean group $E(2)$. From the perspective of an application, Han *et al.* [16] and Gupta *et al.* [15] extract rotation-equivariant feature maps for oriented object detection and visual tracking, respectively. We leverage $E(2)$ -equivariant CNNs [44] as a building block of our network to perceive consistent symmetry patterns across multiple orientations.

2.2. Symmetry detection

Symmetry detection deals with different kinds of symmetric patterns such as reflection axis [1, 11, 13, 14, 20, 32, 37, 38, 41–43], rotation center [13, 20, 22, 23, 32, 37, 42], and translation lattice [17, 24, 28, 30, 35, 42, 48].

Sparse prediction. Rotation and translation symmetries are often formulated as periodic signals and detected by autocorrelation [24, 28] in the spatial domain, spectral density [22, 23] and angular correlation [20] in the frequency domain. Meanwhile, there is a consistent need for an affine-invariant or equivariant feature descriptor in detecting symmetries, as matching the local descriptors is the most common solution. Loy and Eklundh [32] use SIFT [31] descriptors and normalize each descriptor by its dominant orientation. Cho and Lee [3] also use SIFT [31] to match feature pairs and detect symmetry by discovering the clusters of the nearby matches. Contour and edge features [1, 37, 39, 42, 43] are also useful for determining the boundary of the symmetric object. Lee and Liu [23] propose to solve affine-skewed rotation symmetry group detection by rectifying the skewed patterns. In this paper, we tackle the task by data-driven approach with our proposed dataset. Also, we use the dihedral group to interpret the symmetries as in many symmetry detection literatures [22, 23].

Dense prediction. Recently proposed methods [11, 13, 38, 41] predict pixel-wise symmetry scores. Fukushima and Kikuchi [11] build a neural network to extract edges from images and detect reflection symmetry. Tsogkas *et al.* [41] construct a bag of features using histogram, color, and texture for each pixel and adopt multiple instance learning when training the model. Gnutti *et al.* [14] take two

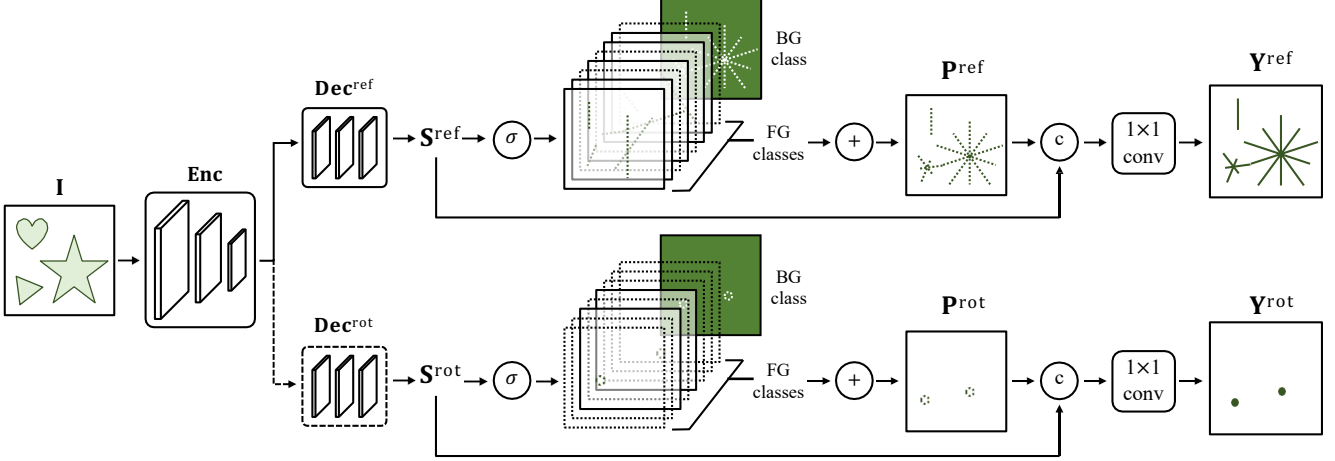


Figure 2. Illustration of the proposed symmetry detection network, EquiSym. After an input image \mathbf{I} passed a group-equivariant encoder Enc , group-equivariant decoders Dec^{ref} and Dec^{rot} predict intermediate predictions \mathbf{S}^{ref} and \mathbf{S}^{rot} for rotation and reflection, respectively. Auxillary tasks for the rotation and reflection symmetry are the order(N) of the rotation fold and the orientation of the reflection axis. The foreground logits are pooled to \mathbf{P}^{ref} and \mathbf{P}^{rot} and stacked with the scores \mathbf{S}^{ref} and \mathbf{S}^{rot} , respectively. The final score \mathbf{Y}^{ref} and \mathbf{Y}^{rot} for the rotation center and the reflection axis are predicted using a group-equivariant 1×1 convolution. For details, see Sec. 3.

stages of computing the symmetry score for each pixel using patch-wise correlation and validating the obtained candidate axes using gradient direction and magnitude. Funk and Liu [13] are the first to adopt deep CNNs for detecting reflection and rotation symmetries. Seo *et al.* [38] propose a polar self-similarity descriptor for better rotation and reflection invariance. A specially designed Polar Matching Convolution (PMC) performs region-wise feature matching to compute the symmetry score, but the model relies heavily on the CNNs. To discover consistent symmetry patterns *w.r.t.* geometric transformations of rotation and reflection, we deploy group equivariant neural networks in our symmetry detection model.

3. Proposed Method

The symmetry patterns appearing in an image are invariant to the 2D rigid transformations of the image. The detected symmetry patterns of a transformed input image should be the same with the transformed detection results using the original input image. In other words, the reflection and rotation equivariance is crucial for a symmetry detection model. To this end, we propose a unified framework for detecting the reflection and rotation symmetries via equivariant learning, EquiSym. The overall pipeline is briefly illustrated in Fig. 2. Given an input image, a shared encoder Enc and decoders Dec^{ref} and Dec^{rot} are applied for detecting reflection and rotation symmetries, respectively. We also perform auxiliary pixel-wise classification tasks, one for the orientation of the reflection axis and the other for the order of the rotation symmetry. The intermediate logits and the corresponding probabilities of the subtasks are denoted by \mathbf{S} and \mathbf{P} , respectively. The logits \mathbf{S} are integrated with

the sum of the foreground probabilities \mathbf{P} to compute the final score map \mathbf{Y} using a 1×1 group-equivariant convolution layer. The following sections cover the preliminaries and the proposed symmetry detection network.

3.1. Preliminaries

Group and equivariance. A group (G, \cdot) is a set G with a binary operation \cdot , where its elements are closed under the operation. A group has unique identity element and inverse element and also satisfies the associativity. Equivariance of a map $f : X \rightarrow Y$ is formalized using a group G and two G -sets X and Y , where G -set is a mathematical object consisting of a set S and a group action of G on S . A map f is said to be equivariant iff

$$f(g \cdot x) = g \cdot f(x), \quad (1)$$

for all $x \in X$ and all $g \in G$. In 2D image domain, we focus on an euclidean group $E(2)$, which is a group of plane \mathbb{R}^2 isometries of translations, rotations, and reflections.

$E(2)$ -steerable feature field. The affine transformations of the 2D or 3D coordinates are easily done by matrix multiplications. Unlike low-dimensional feature vectors, the reflection or rotation transformation of the high-dimensional feature vectors are non-trivial. The first step to build a *steerable* convolution is to define a *steerable* feature field $f : \mathbb{R}^2 \rightarrow \mathbb{R}^c$ that maps a feature vector $f(x) \in \mathbb{R}^c$ with each point x of a base plane. Given a group G , a group representation $\rho : G \rightarrow GL(\mathbb{R}^c)$ specifies the transformation law for shuffling the c channels of each feature vector, where a general linear group GL is the set of $c \times c$ invertible matrices. Thus, applying transformation on a feature map

not only moves the target vectors to the new positions but also shuffles each vector via $\rho(g)$ where $g \in G$. The group representations of E(2) group are presented in [44].

E(2)-equivariant steerable convolution. To preserve the transformation law of the steerable feature spaces in CNNs, equivariance under the group actions is required for each network layer. Convolutions with the restricted *G-steerable kernels* [44] provide an equivariant linear mapping between the steerable feature spaces. The input and output of the *G-steerable* layers are the feature fields with their group representations $\rho_{\text{in}}(g) \in \mathbb{R}^{c_{\text{in}} \times c_{\text{in}}}$ and $\rho_{\text{out}}(g) \in \mathbb{R}^{c_{\text{out}} \times c_{\text{out}}}$, where a group element g is specified. A kernel $k : \mathbb{R}^2 \rightarrow \mathbb{R}^{c_{\text{out}} \times c_{\text{in}}}$ that transforms under ρ_{in} and ρ_{out} becomes *G-steerable* when satisfying a kernel constraint of

$$k(gx) = \rho_{\text{out}}(g)k(x)\rho_{\text{in}}(g^{-1}), \quad (2)$$

for every $g \in G$ given $x \in \mathbb{R}^2$. E(2)-equivariant CNNs solve this constraint to get a basis of the steerable kernels and compute the convolutional weights, which results in the smaller learnable parameters compared to the CNNs.

3.2. Symmetry detection network

Reflection and rotation equivariant modules. Since we aim to establish both reflection and rotation symmetry, we employ an E(2)-equivariant CNNs of dihedral group D_N , which contains N discrete rotations by angles multiples of $\frac{2\pi}{N}$ and reflections. The encoder Enc consists of an E(2)-equivariant [16,44] ResNet [18] and an Atrous Spatial Pyramid Pooling (ASPP) [2] module. The decoder Dec is a 3-layer convolution module. The encoder and decoder designs follow [38] except that all convolution layers are substituted by the E(2)-equivariant convolution layers. During the forward computation, the feature fields are transformed into the predefined fields of the group D_N . For the predictions \mathbf{S}^{ref} and \mathbf{S}^{rot} , the feature fields of the decoders Dec^{ref} and Dec^{rot} are transformed back to the scalar fields.

Auxiliary classification. Instead of a direct regression of the symmetry score maps [13,38], we perform relevant sub-tasks that can lead to the final prediction. The proposed auxiliary tasks are the pixel-wise classification of the orientation (angle) of the reflection axis and the number of rotation folds. For simplicity, we assign the orientation into N^{ref} bins dividing 180 degrees. The ground-truth orientation $\mathbf{S}_{\text{gt}}^{\text{ref}}$ is then quantized to be one-hot. The auxiliary rotation label $\mathbf{S}_{\text{gt}}^{\text{rot}}$ is the order of the rotational symmetry, which is annotated with a positive integer in the case of the discrete rotational symmetry. We allocate '0' to the continuous group since its ground-truth order is infinite. The size of the unique set of the rotation orders in the dataset is denoted as N^{rot} . Meanwhile, we add background classes for

the pixels that are neither the axis nor the center. Therefore, the classifiers predict the scores of channel of $N^{\text{ref}} + 1$ and $N^{\text{rot}} + 1$. The classification logit $\mathbf{S} \in \mathbb{R}^{H \times W \times (N+1)}$ is obtained by

$$\mathbf{S} = \text{Dec}(\text{Enc}(\mathbf{I})). \quad (3)$$

The encoder Enc is shared while the decoders Dec^{ref} and Dec^{rot} are task-specific. Note that we set the group orientation N as the same as N^{ref} to further exploit the equivariance of the equivariant networks.

Symmetry detection. The predicted score maps of the symmetry axes of the corresponding orientation are present in the N^{ref} foreground channels of the estimated orientation \mathbf{S}^{ref} , whereas the background channel contains the background pixels. Similar to reflection, the N^{rot} foreground channels are the score maps of the rotation centers with that number of folds. We aggregate the sum of the foreground logits $\mathbf{P} \in \mathbb{R}^{H \times W \times 1}$ and the intermediate prediction $\mathbf{S} \in \mathbb{R}^{H \times W \times (N+1)}$ to compute the final prediction $\mathbf{Y} \in \mathbb{R}^{H \times W}$ as

$$\mathbf{P}_{h,w} = \sum_{k=1}^N \frac{\exp(\mathbf{S}_{h,w,k})}{\sum_c \exp(\mathbf{S}_{h,w,c})}, \quad (4)$$

$$\mathbf{Y} = \text{conv}_G([\mathbf{P} || \mathbf{S}]). \quad (5)$$

Note that $||$ denotes the concatenation operation along the final channel dimension.

3.3. Training objective

To train EquiSym, we optimize a combination of two loss terms for localization and classification. Following [38], we adopt the focal loss [25] as the localization loss \mathcal{L}_{loc} for both reflection and rotation score maps. The classification loss \mathcal{L}_{cls} for the intermediate predictions is the cross-entropy loss. The final objective \mathcal{L} are expressed as

$$\mathcal{L}_{\text{loc}} = \mathcal{L}_{\text{focal}}(\mathbf{Y}, \mathbf{Y}_{\text{gt}}) \quad (6)$$

$$\mathcal{L}_{\text{cls}} = \mathcal{L}_{\text{ce}}(\mathbf{S}, \mathbf{S}_{\text{gt}}) \quad (7)$$

$$\mathcal{L} = \mathcal{L}_{\text{loc}} + \mathcal{L}_{\text{cls}}, \quad (8)$$

The network for reflection and rotation symmetry detection are denoted as EquiSym-ref and EquiSym-rot, respectively. To alleviate the class imbalance issue, the loss of the background class is weighted with w for \mathcal{L}_{cls} . Note that the focal loss alleviates the class imbalance issue for \mathcal{L}_{loc} .

4. New Symmetry Dataset (DENDI)

We present a new dataset for symmetry detection named *DENSE and DIVERSE symmetry dataset* (DENDI) in the following.

dataset	Ref. split	Rot. split	complexity
SDRW [27]	51 / - / 70	10 / - / 66	low
NYU [4]	239 / - / -	-	low
SymCOCO [12]	250 / - / -	250 / - / -	high
DSW [12]	-	200 / - / -	low
BPS* [13]	959 / - / 240	846 / - / 211	high
LDRS [38]	1,110 / 127 / 240	-	high
DENDI	1,750 / 374 / 369	1,459 / 313 / 307	high

* Not available online.

Table 1. Comparison of the symmetry detection datasets.

4.1. Motivation

Limitations in existing datasets. The early reflection symmetry datasets [4, 27] contain small number of images with few reflection axis and rotation center. Recently proposed BPS [13] and LDRS [38] are large enough to train deep architectures, but the reflection axes still lack of diversity in terms of the length and orientation. For example, objects with multiple symmetry axes are often annotated only by a single dominant axis. Furthermore, no existing reflection symmetry datasets take into account the continuous symmetry group, such as a circle with an infinite number of reflection symmetry axes. For the rotation symmetry, the annotations of BPS [13] are limited to the rotation centers while the pioneer unsupervised methods [22, 23] tackle the rotation folds also.

Proposed dataset. To address the concerns mentioned above, we present a new dataset for reflection and rotation symmetry detection that includes a wide range of geometries. We integrate 239 images of NYU [4] and 181 images of SDRW [27], and collect 2,080 images from COCO [26] dataset. Both reflection and rotation annotations are labeled for each image, and we remove images without any labels. As a result, DENDI contains 2,493 and 2,079 images for reflection and rotation split, respectively. The sizes of the symmetry detection datasets are compared in Tab. 1. To add reflection axes with diverse length and orientation, we annotate objects in common shapes such as circle, ellipse, and polygons, as well as the part-level symmetries. Also, the annotators are encouraged to exhaustively label the symmetries for each object, including the non-dominant ones, *e.g.* the diagonals of a square. For the reflection symmetry of a continuous symmetry group, we annotate with ellipse-shaped masks to represent an infinite number of line axes. For the rotation symmetry, we additionally collect the number of rotation folds for each rotation center. As a result, the annotations in DENDI are denser and more diverse compared to the ones in the existing datasets.

4.2. Annotation

Reflection symmetry. A reflection symmetry axis is defined as a line formed by two points following [4, 12, 13, 27,

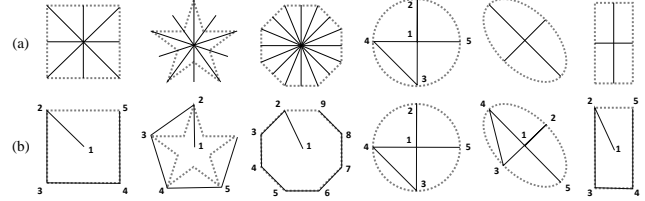


Figure 3. Illustration of the generic shapes and their annotations. (a) and (b) indicate the annotation rules of the reflection and rotation symmetry, respectively. For details, see Sec. 4.2



Figure 4. The images and labels of the objects with generic shapes. (a) and (b) indicate the annotations of the reflection and rotation symmetry, respectively. Best viewed in the electronic version.

[38]. In contrast to the existing datasets, we now account for circular objects. A circular object, which is equivalent to a filled circle, has an infinite number of reflection symmetry axes through its center. We propose to annotate the circular objects with 5 connected points, resembling the shape of the Arabic number '4'. We draw a '4'-shaped annotation from the circle's center towards the circle's boundary in the up, down, left, and right directions. The annotation rules and visualizations of the reflection symmetry for generic shapes are visualized in Fig. 3 (a) and Fig. 4 (a), respectively.

Rotation symmetry. We collect the rotation center coordinate, the object's boundary, and the number of folds (N) for each object. A circular object is in a continuous rotation group with infinite folds. Thus, we set the N as 0 for simplicity. We categorize the objects into ellipses and polygons based on the their shape. Circular or elliptical objects are marked with '4'-shaped annotation. We annotate the polygons of V vertices with (V+1) consecutive points. From the center of the object, we take the vertex closest to 12 o'clock as the 2nd point and link the vertices of a convex polygon counter clockwise. Note that the number of the vertices (V) and the number of the folds (N) not always match. The annotation rules and visualizations of the rotation symmetry for generic shapes are visualized in Fig. 3 (b) and Fig. 4 (b), respectively.

4.3. Statistics

Reflection symmetry. Histograms for scale and orientation of the reflection axes are presented in Fig. 5 (a) and (b). For scale, we measure the length of each line and nor-

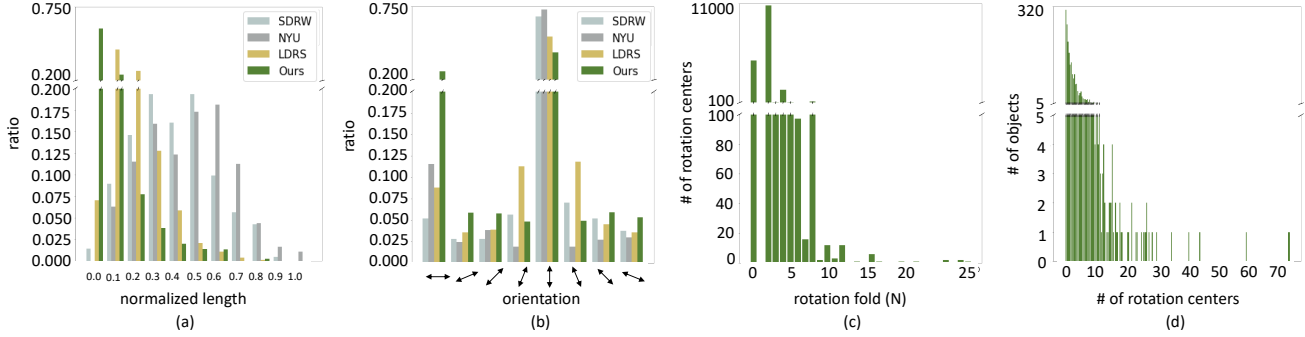


Figure 5. Statistical analysis of DENDI. (a) and (b) represent the reflection symmetry dataset while (c) and (d) indicate the rotation symmetry dataset. In specific, (a) and (b) are histograms for scale and orientation of the reflection axes, (c) and (d) represent the histograms of the rotation fold and the number of the rotation centers.

malize it with the length of the image diagonal. With two points on the line annotation, we can also compute its orientation (tangents). The y-axis of Fig. 5 (a) and (b) is the ratio of the number of the axes over the total number of the axes. The number of axes increases when the length of each axis decreases, reflecting the characteristics of the real-world environment. In addition, it can be interpreted that part-level symmetry is densely annotated compared to other datasets. Note that our dataset ranks first at all orientations except for the three orientations that are closest to the vertical direction. Despite the fact that LDRS [38] is also based on COCO [26], the distribution of the orientation of the axis is more diverse in our case due to the fact that we particularly request the annotators not to omit the non-dominant axes.

Rotation symmetry. Histograms of the rotation fold and the number of the rotation centers are illustrated in Fig. 5 (c) and (d). The three most common folds in the dataset are 2, 0 (continuous), and 4. The result is predictable as there are many rectangles, circles, and squares in the image. Note that the dataset contains a notable number of objects of fold 8, which are mostly the ‘STOP’ signs in the road. The complexity of our rotation symmetry dataset is high, as shown in Fig. 5 (d). Even if we clip a few exceptions, a lot of rotation centers are marked in the images.

5. Experiments

5.1. Experimental settings

Datasets. We use SDRW [27], LDRS [38], and DENDI to evaluate the reflection symmetry detection model. We follow the training and evaluation settings of PMCNet [38] in Tab. a.1 for additional use of NYU [4] and the synthesized images. For rotation, we only use DENDI which contains the images of SDRW [27] rotation dataset.

Evaluation. To evaluate EquiSym, we use F1-score computed with the precision and recall as $\frac{2 \times \text{prec} \times \text{rec}}{\text{prec} + \text{rec}}$. A con-

vention [13, 38, 41] for measuring the score of the output score map is morphological thinning [34] and an off-the-shelf pixel-matching algorithm to compare with the ground-truth lines, which are also pixel-width. In contrast to the existing datasets, we take the circular objects into account, which are annotated with filled circles. As DENDI contains annotations of filled circles, the thinning operation shrinks to a single-pixel dot. Therefore, it is inevitable to come up with a new way to evaluate the predicted score maps. We dilate the ground-truth score maps of reflection and rotation symmetries with a maximum distance of 5 pixels following [13] where they enlarge the ground-truth dots to 5-pixel radius circles around the rotation centers. The predicted score map is also dilated to make the result of the ground-truth itself become 1. The true-positives are then computed by pixel-wise comparisons.

Implementation details. As a backbone network, we adopt the ReResNet implementation of ReDet [16] of depth 50, which is based on PyTorch [36] and e2cnn [44]. The number of the layers and the structure are the same as the vanilla ResNet [18]. We pretrain the ReResNet50 for the image classification task in imagenet-1k [10] following the procedures in [16]. For the symmetry group to initialize the equivariant networks, we use a dihedral group of eight orientations (D_8). To provide multi-scale contexts, we also deploy the Atrous Spatial Pyramid Pooling module (ASPP) [2] which we re-implemented the module by replacing all the vanilla convolutions to E(2)-equivariant convolutions [44]. The number of the classes N^{ref} is 8 and N^{rot} is 21. We train EquiSym for 100 epochs with an initial learning rate of 0.001 using the Adam [21] optimizer with a batch size of 32. For details, refer to the supplementary material.

5.2. Ablation studies

Ablation on the equivariant convolution. We study the effectiveness of the group-equivariant convolution in Tab. 2. With the help of the equivariant convolutions the

model	design choices		F1 score	
	Equiv.	Aux.	Ref.	Rot.
Ref.			55.1	-
	✓		63.1	-
	✓	✓	64.5	-
Rot.			-	17.7
	✓		-	21.2
	✓	✓	-	22.5
joint	✓		62.2	22.1
	✓	✓	58.7	22.5

Table 2. Ablation on the symmetry detection network in reflection, rotation, and joint model on DENDI dataset.

F1 scores of 55.1 and 17.7 increase to 63.1 and 21.2 for EquiSym-ref and EquiSym-rot, respectively.

Ablation on the auxiliary classification. To enhance the intermediate representation, we perform a relevant subtask for each branch of symmetry detection and compare them in Tab. 2. Without extra labels, the reflection-only model achieves the F1 score of 64.5, which is greater than the 63.1 obtained by training solely with the final task. The orientation estimation also requires rotation equivariance, which enhances the intermediate features. Rotation symmetry detection, on the other hand, requires extra annotation for the auxiliary task as the original labels are a set of dots. The rotation invariant subtask of classification of the rotation folds (N) compresses the information of the intermediate feature so that it can increase the F1 score from 21.2 to 22.5.

Ablation on the joint training. We investigate the effect of a joint training of the reflection and rotation symmetries in Tab. 2. When training EquiSym-joint, the loss \mathcal{L} is computed for both reflection and rotation symmetries. Joint symmetry detection network that is trained only with the final task achieves comparable F1 scores of 62.2 and 22.1 for reflection and rotation symmetry, respectively. However, the auxiliary task do not increase the accuracy of the reflection symmetry in joint training scenario. One probable explanation is that only orientation estimation of the reflection symmetry axis requires rotation equivariance while the fold classification only necessitates rotation invariance, resulting in network imbalance.

5.3. Comparison with the state-of-the-art methods

We compare EquiSym with the state-of-the-art methods in Tab. 3 and Tab. a.1. For both reflection and rotation symmetries, our proposed EquiSym achieves the state-of-the-art, showing the effectiveness of the equivariant networks and the auxiliary classification. While PMCNet [38] is re-trained on DENDI for a fair comparison, SymResNet [13]

symmetry	method	F1 score
reflection	SymResNet* [13]	30.7
	PMCNet [†] [38]	52.0
	EquiSym-ref	64.5
rotation	SymResNet* [13]	11.9
	EquiSym-rot	22.5

[†]Re-trained on DENDI.

* Evaluated using the weights provided by the authors.

Table 3. Comparison with the state-of-the-art methods on DENDI.

method	train dataset		test dataset		mixed
	real	synth	SDRW	LDRS	
PMCNet [38]	✓		61.6	34.8	61.2
	✓	✓	68.8	37.3	
EquiSym-ref	✓		67.4	40.9	71.4
	✓	✓	67.1	39.4	

Table 4. Comparison of the reflection symmetry detection methods on the LDRS [38] and SDRW [27]. Note that the real dataset consists of SDRW, LDRS, and NYU [4] dataset.

is compared using the weights provided by the authors as fine-tuning SymResNet [13] degraded the performance.

We follow the configurations of PMCNet [38] for the experiments on SDRW [27] and LDRS [38] in Tab. a.1. The training images consist of real images from SDRW, LDRS, and NYU [4] and the generate synthetic images as in [38]. EquiSym-ref achieves the state-of-the-art on LDRS while the results on SDRW are still comparable. The additional use of synthetic images is not helpful as proposed in [38]. One possible reason is the imbalance in data distribution across splits. To mitigate this issue, we construct a new split denoted as mixed by merging all images and then randomly split the images into *train/val/test* splits with the ratio of 4:1:1. EquiSym-ref outperforms PMCNet in that scenario, as shown in Tab. a.1. All the experiments in Tab. a.1 are evaluated with the legacy scheme.

5.4. Qualitative results

The qualitative results of EquiSym-ref, PMCNet [38], and SymResNet [13] on DENDI-ref *test* are shown in Fig. 6. EquiSym-ref produces dense reflection symmetry score maps compared to other methods, including non-dominant axes such as the diagonal. Furthermore, EquiSym-ref predicts masks of the circular objects accurately even for the challenging samples where the line and circles both exist in the ground-truth. We compare EquiSym-rot with SymResNet on DENDI-rot *test* in Fig. 7. EquiSym-rot is robust in scale and number of rotation centers. EquiSym-rot detects symmetry of polygons as well as circular objects in DENDI-rot, whereas SymResNet mainly detects circular objects.

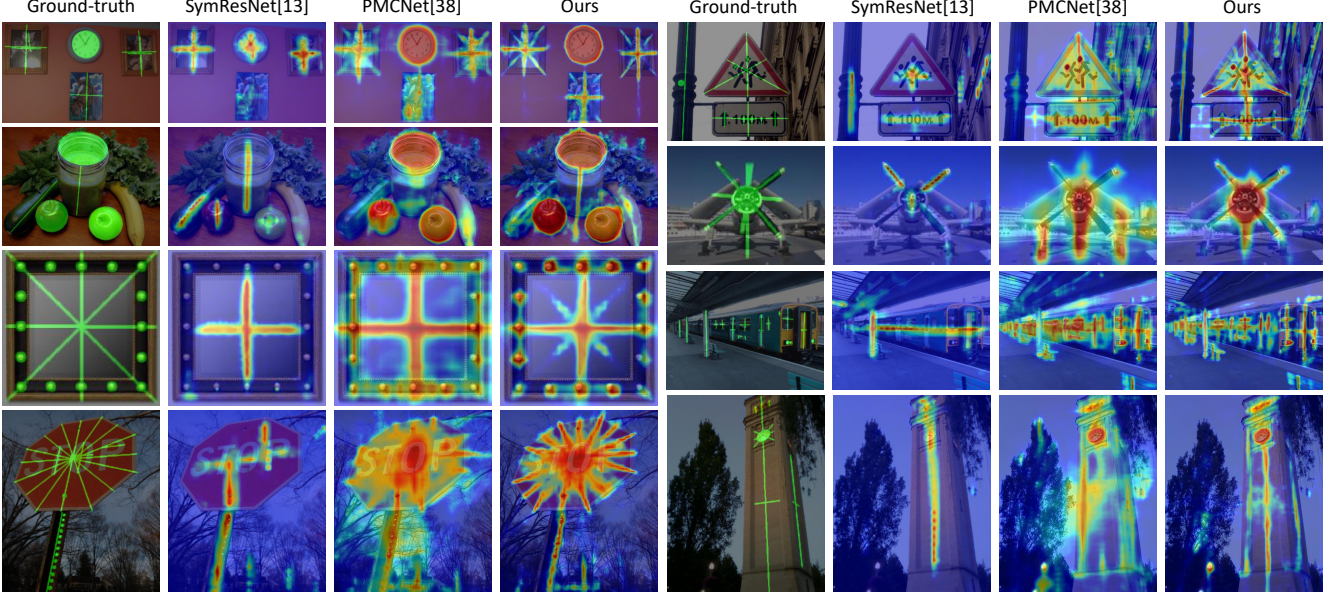


Figure 6. Qualitative results of the reflection symmetry detection on DENDI-ref *test*.

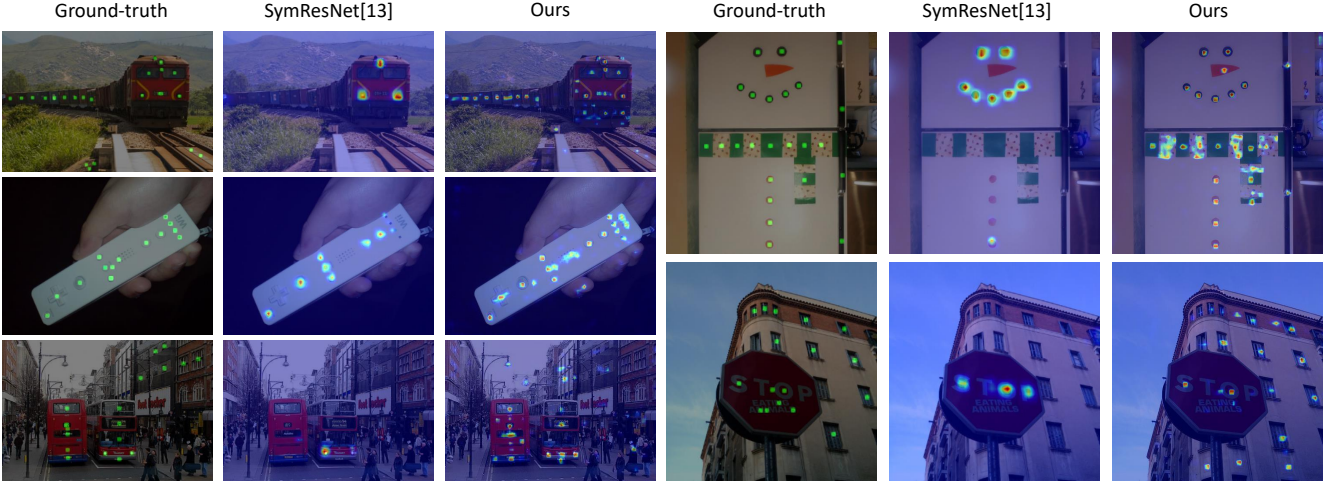


Figure 7. Qualitative results of the rotation symmetry detection on DENDI-rot *test*.

5.5. Limitations

While EquiSym can be jointly trained to produce comparable predictions as the single-branch EquiSym, it has more room for improvement. Especially, the design of the auxiliary task of EquiSym-rot can be explored more to enhance the accuracy of the reflection symmetry detection.

6. Conclusion

In this paper, we have proposed a novel symmetry detection framework, EquiSym, using equivariant learning to obtain group-equivariant and invariant scores for both reflection and rotation symmetries. In addition, we have intro-

duced a new dataset *DENSE and DIVERSE symmetry dataset* (DENDI) for reflection and rotation symmetries. The proposed EquiSym achieves the state-of-the-art on LDRS and DENDI datasets.

Acknowledgements. This work was supported by Samsung Advanced Institute of Technology (SAIT) and also by the NRF grant (NRF-2021R1A2C3012728) and the IITP grant (No.2021-0-02068: AI Innovation Hub, No.2019-0-01906: Artificial Intelligence Graduate School Program at POSTECH) funded by the Korea government (MSIT). We like to thank Yunseon Choi for her contribution to DENDI.

References

- [1] Ibragim R Atadjanov and Seungkyu Lee. Reflection symmetry detection via appearance of structure descriptor. In *Eur. Conf. Comput. Vis.*, pages 3–18. Springer, 2016. 1, 2
- [2] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017. 4, 6, 11
- [3] Minsu Cho and Kyoung Mu Lee. Bilateral symmetry detection via symmetry-growing. In *Brit. Mach. Vis. Conf.*, pages 1–11. Citeseer, 2009. 1, 2
- [4] M. Cicconet, V. Birodkar, M. Lund, M. Werman, and D. Geiger. A convolutional approach to reflection symmetry. <http://arxiv.org/abs/1609.05257>, 2016. New York. 2, 5, 6, 7, 11
- [5] Taco Cohen, Mario Geiger, and Maurice Weiler. A general theory of equivariant cnns on homogeneous spaces. *arXiv preprint arXiv:1811.02017*, 2018. 2
- [6] Taco Cohen, Maurice Weiler, Berkay Kicanaoglu, and Max Welling. Gauge equivariant convolutional networks and the icosahedral cnn. In *International Conference on Machine Learning*, pages 1321–1330. PMLR, 2019. 2
- [7] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016. 2
- [8] Taco S Cohen, Mario Geiger, and Maurice Weiler. Intertwiners between induced representations (with applications to the theory of equivariant neural networks). *arXiv preprint arXiv:1803.10743*, 2018. 2
- [9] Taco S Cohen and Max Welling. Steerable cnns. *arXiv preprint arXiv:1612.08498*, 2016. 2
- [10] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *cvpr*, pages 248–255. Ieee, 2009. 6, 11
- [11] Kunihiro Fukushima and Masayuki Kikuchi. Symmetry axis extraction by a neural network. *Neurocomputing*, 69(16):1827–1836, 2006. 2
- [12] Christopher Funk, Seungkyu Lee, Martin R Oswald, Stavros Tsogkas, Wei Shen, Andrea Cohen, Sven Dickinson, and Yanxi Liu. 2017 iccv challenge: Detecting symmetry in the wild. In *Int. Conf. Comput. Vis. Worksh.*, pages 1692–1701, 2017. 2, 5, 11
- [13] Christopher Funk and Yanxi Liu. Beyond planar symmetry: Modeling human perception of reflection and rotation symmetries in the wild. In *Int. Conf. Comput. Vis.*, pages 793–803, 2017. 1, 2, 3, 4, 5, 6, 7, 11
- [14] Alessandro Gnutti, Fabrizio Guerrini, and Riccardo Leonardi. Combining appearance and gradient information for image symmetry detection. *IEEE Trans. Image Process.*, 2021. 2
- [15] Deepak K Gupta, Devanshu Arya, and Efstratios Gavves. Rotation equivariant siamese networks for tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12362–12371, 2021. 2
- [16] Jiaming Han, Jian Ding, Nan Xue, and Gui-Song Xia. Redet: A rotation-equivariant detector for aerial object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2786–2795, 2021. 2, 4, 6
- [17] James Hays, Marius Leordeanu, Alexei A Efros, and Yanxi Liu. Discovering texture regularity as a higher-order correspondence problem. In *Eur. Conf. Comput. Vis.*, pages 522–535. Springer, 2006. 2
- [18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *cvpr*, pages 770–778, 2016. 4, 6, 11
- [19] Emiel Hoogetboom, Jorn WT Peters, Taco S Cohen, and Max Welling. Hexaconv. *arXiv preprint arXiv:1803.02108*, 2018. 2
- [20] Yosi Keller and Yoel Shkolnisky. A signal processing approach to symmetry detection. *IEEE Trans. Image Process.*, 15(8):2198–2207, 2006. 2
- [21] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *Int. Conf. Learn. Represent.*, 2015. 6
- [22] Seungkyu Lee, Robert T Collins, and Yanxi Liu. Rotation symmetry group detection via frequency analysis of frieze-expansions. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1–8. IEEE, 2008. 2, 5
- [23] Seungkyu Lee and Yanxi Liu. Skewed rotation symmetry group detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 32(9):1659–1672, 2009. 2, 5
- [24] Hsin-Chih Lin, Ling-Ling Wang, and Shi-Nine Yang. Extracting periodicity of a regular texture based on autocorrelation functions. *Pattern recognition letters*, 18(5):433–443, 1997. 2
- [25] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 2980–2988, 2017. 4
- [26] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014. 5, 6
- [27] Jingchen Liu, George Slota, Gang Zheng, Zhaohui Wu, Minwoo Park, Seungkyu Lee, Ingmar Rauschert, and Yanxi Liu. Symmetry detection from realworld images competition 2013: Summary and results. In *IEEE Conf. Comput. Vis. Pattern Recog. Worksh.*, pages 200–205, 2013. 2, 5, 6, 7, 11
- [28] Yanxi Liu, Robert T Collins, and Yanghai Tsin. A computational model for periodic pattern perception based on frieze and wallpaper groups. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(3):354–371, 2004. 2
- [29] Yanxi Liu, Hagit Hel-Or, and Craig S Kaplan. *Computational symmetry in computer vision and computer graphics*. Now publishers Inc, 2010. 1
- [30] Yanxi Liu, Wen-Chieh Lin, and James Hays. Near-regular texture analysis and manipulation. *ACM Transactions on Graphics (TOG)*, 23(3):368–376, 2004. 2
- [31] David G Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.*, 60(2):91–110, 2004. 2

- [32] Gareth Loy and Jan-Olof Eklundh. Detecting symmetry and symmetric constellations of features. In *Eur. Conf. Comput. Vis.*, pages 508–521. Springer, 2006. 1, 2
- [33] Diego Marcos, Michele Volpi, Nikos Komodakis, and Devis Tuia. Rotation equivariant vector field networks. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 5048–5057, 2017. 2
- [34] David R Martin, Charless C Fowlkes, and Jitendra Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5):530–549, 2004. 6
- [35] Minwoo Park, Kyle Broeklehurst, Robert T Collins, and Yanxi Liu. Deformed lattice detection in real-world images using mean-shift belief propagation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(10):1804–1816, 2009. 2
- [36] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 8024–8035. Curran Associates, Inc., 2019. 6, 11
- [37] V Shiv Naga Prasad and Larry S Davis. Detecting rotational symmetries. In *Int. Conf. Comput. Vis.*, volume 2, pages 954–961. IEEE, 2005. 1, 2
- [38] Ahyun Seo, Woohyeon Shim, and Minsu Cho. Learning to discover reflection symmetry via polar matching convolution. In *Int. Conf. Comput. Vis.*, 2021. 1, 2, 3, 4, 5, 6, 7, 11
- [39] Dinggang Shen, Horace HS Ip, and Eam Khwang Teoh. Robust detection of skewed symmetries by combining local and semi-local affine invariants. *Pattern Recognition*, 34(7):1417–1428, 2001. 1, 2
- [40] Ivan Sosnovik, Michał Szmaja, and Arnold Smeulders. Scale-equivariant steerable networks. *arXiv preprint arXiv:1910.11093*, 2019. 2
- [41] Stavros Tsogkas and Iasonas Kokkinos. Learning-based symmetry detection in natural images. In *Eur. Conf. Comput. Vis.*, pages 41–54. Springer, 2012. 2, 6
- [42] Zhaozhong Wang, Lianrui Fu, and YF Li. Unified detection of skewed rotation, reflection and translation symmetries from affine invariant contour features. *Pattern Recognition*, 47(4):1764–1776, 2014. 2
- [43] Zhaozhong Wang, Zesheng Tang, and Xiao Zhang. Reflection symmetry detection using locally affine invariant edge correspondence. *IEEE Trans. Image Process.*, 24(4):1297–1301, 2015. 1, 2
- [44] Maurice Weiler and Gabriele Cesa. General E(2)-Equivariant Steerable CNNs. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019. 2, 4, 6, 11
- [45] Maurice Weiler, Mario Geiger, Max Welling, Wouter Boomsma, and Taco Cohen. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. *arXiv preprint arXiv:1807.02547*, 2018. 2
- [46] Hermann Weyl. Symmetry princeton university press. *Princeton, New Jersey*, page 17, 1952. 1
- [47] Daniel E Worrall, Stephan J Garbin, Daniyar Turmukhambetov, and Gabriel J Brostow. Harmonic networks: Deep translation and rotation equivariance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5028–5037, 2017. 2
- [48] Peng Zhao and Long Quan. Translation symmetry detection in a fronto-parallel view. In *IEEE Conf. Comput. Vis. Pattern Recog.*, pages 1009–1016. IEEE, 2011. 2

Appendix

A.1. EquiSym

The details that are omitted in the main paper are covered in this section. We show the consistency of the evaluation schemes. The implementation details are also shown in the following.

A.1.1. Evaluation scheme

method	train dataset		test dataset		mixed
	real	synth	SDRW	LDRS	
PMCNet [38]	✓	✓	N/A	N/A	46.6
	✓		51.1	33.7	
EquiSym-ref	✓		49.8	36.5	58.0
	✓	✓	49.2	34.9	

Table a.1. Comparison of the reflection symmetry detection methods on the LDRS [38] and SDRW [27]. Note that the real dataset consists of SDRW, LDRS, and NYU [4] dataset and synthetic images generated as in [38].

In the main paper, we propose to use a modified evaluation scheme of blurring the ground truth rather than thinning the prediction. The primary reason is that the thinning process transforms a circular mask prediction into a single dot. The pixel matching algorithm determines whether the predicted lines are close enough to the ground truth lines within a threshold, which becomes equivalent when the ground truth itself is blurred with a radius of a threshold. In practice, we construct a filter of the kernel 11×11 where the weights are set to 1 for a circle of a diameter 11 and 0 otherwise. We convolve the ground-truth heatmap of both symmetries with the filter so the heatmap is dilated to the maximum distance of 5 pixels. The true positives are then computed by pixel-wise comparisons. We re-evaluate the experiments of Tab. 3 of the main paper in Tab. a.1. Note that one experiment from PMCNet [38] is excluded since the trained model is not accessible. As shown in Tab. a.1, the rankings produced by the two evaluation schemes are consistent, while the latter is significantly faster.

A.1.2. Implementation details

Construction of the orientation labels. EquiSym utilizes the intermediate tasks to increase the accuracy of the symmetry detection tasks. In the case of reflection symmetry, the intermediate labels of the orientations of the reflection axes are obtained for free. The angle of the reflection symmetry axis in the form of a straight line can be expressed as a linear combination of the closest one or two angles among the 8 predetermined angles, which is an initial soft label. On the other hand, the circle-shaped symmetry axis has an evenly divided orientation label of 8 segments

determined by the orientation of the line crossing the center. The orientation label is then quantized for training.

ImageNet pretrain. To be consistent with experiments on the vanilla ResNet [18] pre-trained on ImageNet [10], we pretrain the ReResNet50 on ImageNet-1K for the image classification. While ReResNet50 from ReDet is implemented with C_8 group, we use D_8 group instead. Furthermore, we adjust the stride and dilation of each layer in ReResNet50 to obtain a higher resolution feature map with a larger receptive field than the original one, which is a common procedure in the semantic segmentation [2]. The learning rate starts at 0.1 and decreases by 0.1 every 30 epochs, for a total of 100 epochs. We use a batch size of 512. The pretrained ReResNet50 achieves 69.06% top-1 and 87.25% top-5 accuracy on the ImageNet *val*.

Implementation details. Following [38], The hyperparameters α and β of the focal loss are set to 0.95 and 2, respectively. For training, we resize input images so that the maximum length of the width or the height is 417. The background weight w of \mathcal{L}_{cls} is set to 0.01 and 0.001 for reflection and rotation symmetries. We use the PyTorch [36] and e2cnn [44] framework to build our model.

A.2. DENDI

The details of data annotation for the DENDI dataset are described in this section. To identify symmetry, we disregard texture and focus just on the shape of the object. The partial occlusion of boundary of symmetric object is allowed to a fourth of the object boundaries. For both reflection and rotation symmetries, we exhaustively mark all symmetries in an image, including those of parts. This policy ensures that the DENDI contains dense annotations. we present some examples in Fig. a.1 (d) and Fig. a.2 (d).

A.2.1. Reflection symmetry

A reflection symmetry axis is drawn as a line following the previous datasets [4, 12, 13, 27, 38]. The notable examples are shown in Fig. a.1 (a). We annotate all reflection symmetry axes in an object, including non-dominant ones. Different from the existing datasets, we account for a circular object, which has an infinite number of reflection symmetry axes. Instead of an infinite number of symmetry axes to represent a circular object, we use a '4'-shaped label consisting of 5 points which are then converted to a circular mask, as shown in Fig. a.1 (b). Note that a semantically circular object that seems to be an ellipse due to viewpoint variants is also annotated with a '4'-shaped label, as shown in the first two rows in Fig. a.1 (c). Likewise, a skewed regular polygon due to perspective variations has the same reflection axis as a non-skewed regular polygon, as shown



Figure a.1. Illustration of the examples in the reflection symmetry dataset. The samples with (a) multiple symmetry axes, (b) circular objects, (c) skewed objects, and (d) dense symmetry axes are shown in the figure. Green lines indicate the reflection axes and the yellow lines indicate the '4'-shaped reflection circle annotation. The reflection-circle annotations are then converted to masks.

in the last two rows in Fig. a.1 (c), *e.g.*, a regular STOP sign and a skewed STOP sign that are both semantically regular octagon have eight reflection symmetry axes.

Furthermore, we annotate symmetry in characters such as A, B, C, D, E, H, I, K, M, O, T, U, V, W, X, and Y, as well as the numbers 0, 1, 2, 5, and 8, except for those that are too thin or indistinct. We also annotate symmetry in the D-shaped part of characters, such as P and R. If multiple symmetry axes are overlapped, only the longest one is saved.

A.2.2. Rotation symmetry

For each object with rotation symmetry, we collect the coordinate of the rotation center, the boundary of the object, and the number of the rotation folds (N). We again employ

the '4'-shaped labels to denote circular or elliptical objects as shown in Fig. a.2 (a). The semantically circular object also features an infinite number of rotation folds, indicated as 0 for simplicity, in addition to the '4'-shaped labels. Similar to reflection symmetry, an semantically circular object with elliptical shape due to viewpoint variants has a rotation fold of 2, *e.g.*, the third and fourth rows in Fig. a.2 (a). The minor axis takes precedence over the major axis when drawing '4'-shaped labels for elliptical objects. The rotation fold of a circular object, in particular, can be greater than 2 if the object contains cyclic symmetry. In the case of a non-circular object with rotation symmetry such as Fig. a.2 (c), we draw a convex polygon starting from the center of the object and following convex vertices. The vertex nearest to 12 o'clock takes priority among the convex vertices. Like-

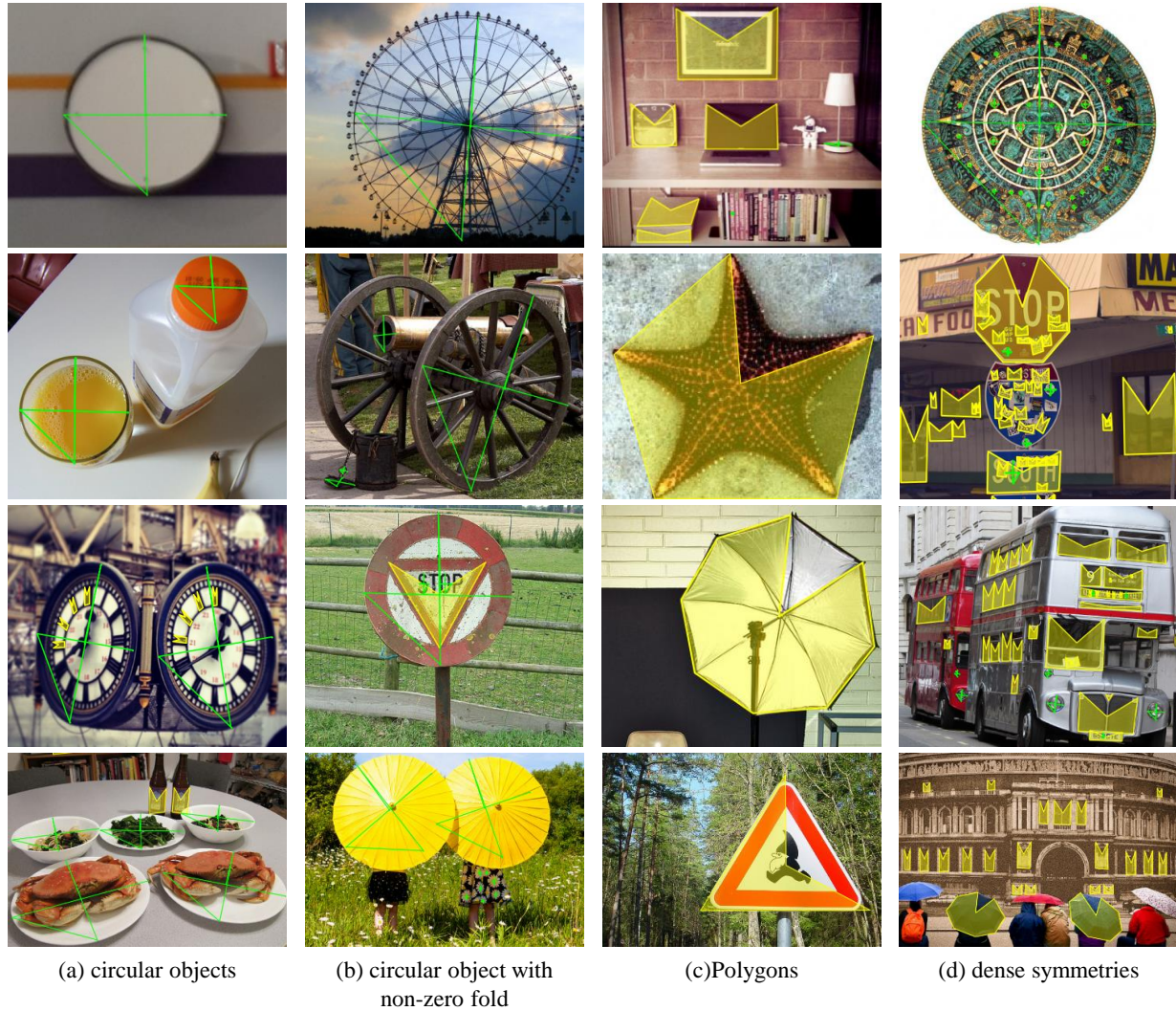


Figure a.2. Illustration of the examples in the rotation symmetry dataset. The samples with (a) circular objects, (b) circular objects with folds larger than 2, (c) polygons, and (d) dense symmetries are shown in the figure. Green lines indicate the circular annotations and the yellow polygons indicate the polygon-type annotations. Only the center coordinates are used for evaluation.

wise, in the reflection symmetry dataset, symmetry in characters such as H, I, N, O, S, X, and Z, as well as the numbers 0 and 8, are taken into account.