Relaxed Rotational Equivariance via G-Biases in Vision

Zhiqiang Wu[1*], Yingjie Liu[1], Licheng Sun[1*], Jian Yang[2], Hanlin Dong[1], Shing-Ho J. Lin[3], Xuan Tang[4], Jinpeng Mi[5], Bo Jin[6], Xian Wei[1†]

[1]Software Engineering Institute, East China Normal University
[2]School of Geospatial Information, Information Engineering University
[3]School of Artificial Intelligence, University of Chinese Academy of Sciences
[4]School of Communication and Electronic Engineering, East China Normal University
[5]Institute of Machine Intelligence, University of Shanghai for Science and Technology
[6] School of Computer Science and Technology, Tongji University
{51265902095, lcsun}@stu.ecnu.edu.cn, xwei@tum.de

Abstract

Group Equivariant Convolution (GConv) can capture rotational equivariance from original data. It assumes uniform and strict rotational equivariance across all features as the transformations under the specific group. However, the presentation or distribution of real-world data rarely conforms to strict rotational equivariance, commonly referred to as Rotational Symmetry-Breaking (RSB) in the system or dataset, making GConv unable to adapt effectively to this phenomenon. Motivated by this, we propose a simple but highly effective method to address this problem, which utilizes a set of learnable biases called G-Biases under the group order to break strict group constraints and then achieve a Relaxed Rotational Equivariant Convolution (RREConv). To validate the efficiency of RREConv, we conduct extensive ablation experiments on the discrete rotational group $\mathcal{C}_n$. Experiments demonstrate that the proposed RREConv-based methods achieve excellent performance compared to existing GConv-based methods in both classification and 2D object detection tasks on the natural image datasets.

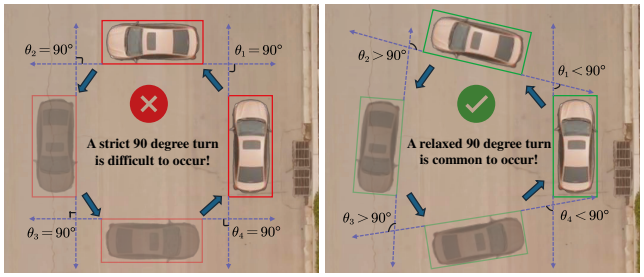
Code — https://github.com/wuer5/rrenet
Extended version — https://arxiv.org/abs/2408.12454

Introduction

Symmetry prior, such as equivariance, plays a vital role in deep learning (Bogatskiy et al. 2022; He et al. 2021; Esteves 2020; Ravanbakhsh, Schneider, and Poczos 2017). Given the assumption of perfect symmetry in data, recent works on equivariant networks are constrained to operate as strict equivariant or invariant functions. These works have been shown to learn potential equivariance or symmetry information without additional data, achieving excellent results with improved efficiency and generalization ability (Cohen and Welling 2016, 2017; Kondor and Trivedi 2018; Ghosh et al. 2022; Kaba et al. 2023; Kaba and Ravanbakhsh 2023).

However, real-world physical systems seldom conform to perfect or strict symmetry due to the natural laws of absolute motion in the world, a phenomenon commonly termed Symmetry-Breaking (Wang, Walters, and Yu 2022; Barone and Theophilou 2008; Wang et al. 2024; Vernizzi and Wheater 2002; Ghosh et al. 2022; Kaba and Ravanbakhsh 2023). In fact, Symmetry-Breaking is a broad concept that involves different definitions of objects or systems. Usually, one definition can refer to breaking the symmetry of an object or system itself. Another definition can be the motion of an object or system that cannot precisely conform to the rules of a strict symmetry group (e.g., the typical rotational group $\mathcal{C}_n$), thereby breaking its strict symmetry state.

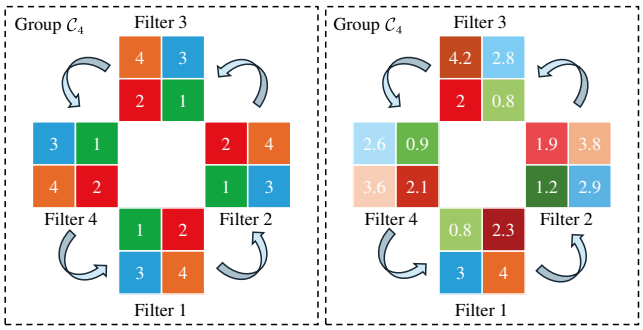
This paper specifically focuses on the second definition of Symmetry-Breaking in the following context. We observe that this type of Symmetry-Breaking often occurs within an object's potential motion in 2D vision. A common example of breaking the strict symmetry state on the group $\mathcal{C}_4$ within a car can be seen in Figure 1. In (a), a car turns by 90 degrees, which can be challenging to achieve precisely in practical situations. Conversely, in (b), a car turns by more or less than 90 degrees, often with a slightly randomized angle change,



(a) Less common scenarios (b) More common scenarios

Figure 1: (a) A car turning right at an angle of exactly 90 degrees denotes the strict adherence to the motion rules on the group $\mathcal{C}_4$. (b) Another car turning right at an angle of approximately 90 degrees represents a deviation from strict rotational symmetry on the group $\mathcal{C}_4$, leading to Rotational Symmetry-Breaking (RSB) within a car's motion. Note that the figure emphasizes the symmetry of an object's potential motion, not the symmetry of an object itself on the group $\mathcal{C}_4$.

*These authors contributed equally.
†Corresponding author.



(a) GConv Filters (SRE) (b) RREConv Filters (RRE)

Figure 2: The 2×2 filters between Strict Rotational Equivariance (SRE) and Relaxed Rotational Equivariance (RRE) on the group $\mathcal{C}_4$. Filters 1 to 4 in Figure (a) have the same values in four directions, whereas Filters 1 to 4 in Figure (b) have slightly different values in four directions.

reflecting a more realistic depiction of a car's turning motion in real-world settings. Strict rotational symmetry constraints hinder equivariant networks from effectively modelling real-world scenarios and visual perception.

Following the second definition above, we mainly concentrate on a common case, i.e., Rotational Symmetry-Breaking (RSB) based on GConv (Cohen and Welling 2016) on the discrete rotational group $\mathcal{C}_n$ in vision domain. By rethinking the construction process and principles of GConv, we find that under a strict group G, each G-transformation convolution filter shares the same copy value, differing only in their positions, which is the key to achieving GConv's strict equivariance. For example, GConv cannot effectively capture the equivariance of objects or scenes with rotations other than 90, 180, and 270 degrees on the group $\mathcal{C}_4$. Therefore, we aim to relax the strict G-transformation convolution filter value-sharing problem to adapt to RSB.

Inspired by convolution biases, we innovatively introduce a set of learnable biases called G-Biases to add to the G-transformation convolution filters. In this paper, we refer to GConv with G-Biases as Relaxed Rotational Equivariant Convolution (RREConv) on the group $\mathcal{C}_n$. The rotational equivariance of GConv is called Strict Rotational Equivariance (SRE), and of RREConv is called Relaxed Rotational Equivariance (RRE). The difference between GConv and RREConv filters can be seen in Figure 2. Note that G-Biases are learnable parameters that can update end-to-end based on the characteristics or distribution of the dataset.

The main contributions are as follows:

- To the best of our knowledge, we are the first to explore RSB within an object's potential motion in vision.
- We propose a simple yet efficient method to address RSB based on the existing GConv.
- The proposed RREConv enhances the performance of GConv-based models with fewer additional parameters and is easily integrated as a plug-and-play module across various GConv-based models.

Related Works

Strict Rotational Equivariance (SRE). SRE is critical for neural networks capturing rotational equivariance from original data, especially in computer vision tasks involving 2D images and 3D objects (Marcos, Volpi, and Tuia 2016). As demonstrated by pioneering work (Cohen and Welling 2016), introducing group equivariance in traditional CNNs led to the development of novel G-CNNs. Foundational networks such as (Li et al. 2018; Marcos et al. 2017; Chidester, Do, and Ma 2018; Weiler and Cesa 2019), have achieved notable success in exploring the rotational equivariance of images. In (Veeling et al. 2018; Müller et al. 2021), the application of rotational equivariance in medicine has achieved excellent performance due to the frequent presence of rotational equivariance in medical data. Moreover, rotational equivariance is also prevalent in 2D object detection. For example, ReDet (Han et al. 2021) introduces rotational equivariance in detecting aerial objects. MORE-Net (Zhu et al. 2022) proposes the multi-oriented ground objects detector to extract rotation-invariant semantic representations. In addition, EquiSym (Seo et al. 2022) outputs group equivariant score maps for rotational centres through end-to-end rotational equivariant feature maps. Rotational equivariance has also seen significant advancements in 3D vision. In (Weiler et al. 2018; Fuchs et al. 2020), 3D Steerable CNNs and SE(3)-Transformers are proposed to explore 3D rotational equivariance. In 3D object detection, 3D equivariance remains crucial. DuEqNet (Wang et al. 2023) focuses on introducing 3D rotation group equivariance constraints in 3D object detection tasks, highlighting the importance of rotational equivariant features in enhancing the robustness of 3D detection models. ProEqBEV (Liu et al. 2024) demonstrates the effectiveness of rotational equivariant BEV features. These methods preserve the structural integrity of the learned features under rotation, enhancing the robustness of the network to such strict rotational transformations.

Relaxed Rotational Equivariance (RRE). However, the methods of SRE mentioned above make them poorly suited for handling RSB in both 2D and 3D fields, as they assume that the rotational equivariance exhibits perfect rotational symmetry, which is a rare case in real-world scenarios (Wu, Hu, and Kong 2015; Dieleman, De Fauw, and Kavukcuoglu 2016; Kavukcuoglu et al. 2009). In contrast, explorations into RRE reveal a significant gap in current research. Although SRE models are effective under ideal conditions, they fall short when dealing with the more common, imperfect symmetries in the real world. In some theoretical works (Kaba and Ravanbakhsh 2023; Kaba et al. 2023), they meticulously analyze the importance and potential application scenarios of relaxed equivariance. As highlighted in (Romero and Lohit 2022), partial rotational equivariance is more effective in representing real-world data than full rotational equivariance. Similarly, (van der Ouderaa, Romero, and van der Wilk 2022) emphasizes that the constraints in equivariance can be too restrictive. In addition, (Wang et al. 2024) explores the relaxed equivariance of physical systems. These methods confirm that relaxed equivariance (e.g., especially RRE) is more suitable for real-world scenarios.

Rotational Symmetry and Symmetry-Breaking. Rotational symmetry (Li, Nagano, and Terashi 2024), a broad concept in natural systems, refers to objects or patterns that retain their appearance when rotated at specific angles (Barone and Theophilou 2008). Although prevalent in theoretical models, this ideal or pristine form of symmetry is often disrupted in real-world scenarios, a phenomenon referred to as RSB (Vernizzi and Wheater 2002). Recent progress in understanding these deviations has spurred the creation of new methodologies aimed at tackling these challenges. For instance, methods suggested by (Desai, Nachman, and Thaler 2022) propose techniques to integrate Symmetry-Breaking elements into models, enhancing their performance in data mining and analysis tasks. However, most researchers focus on strict rotational symmetry and often overlook the phenomenon of RSB. Therefore, this paper primarily focuses on addressing this phenomenon.

Preliminary

Definition of Strict Equivariance. Assume that the input group representation φ_X of G acts on X and the output group representation φ_Y of G acts on Y. A learnable function $f_{\text{strict}} : X \to Y$ satisfies Strict Equivariance if

$$f_{\text{strict}}(\varphi_X(g)(\mathbf{x})) = \varphi_Y(g) f_{\text{strict}}(\mathbf{x}), \tag{1}$$

where all $\mathbf{x} \in X$ and $g \in G$.

Definition of ε-Relaxed or ε-Approximate Equivariance. Consistent with the above definition, a learnable function $f_{\text{relaxed}} : X \to Y$ satisfies Relaxed Equivariance if

$$\|f_{\text{relaxed}}(\varphi_X(g)(\mathbf{x})) - \varphi_Y(g) f_{\text{relaxed}}(\mathbf{x})\| \leq \varepsilon, \tag{2}$$

where all $\mathbf{x} \in X$ and $g \in G$. The upper bound ε is usually a small number, with a relatively larger value indicating the greater level of relaxation and a relatively smaller value showing the stronger level of equivariance. Especially when $\varepsilon = 0$, we have $f_{\text{relaxed}} = f_{\text{strict}}$. Note that ε is determined by the level of Symmetry-Breaking of the dataset or system, which is an implicit constant.

Strict Equivariant Network. Given a set of strict equivariant functions $\{f_i\}$, a strict equivariant network can be the composition function of $\{f_i\}$. Assume f_1 and f_2 satisfy strict equivariance and $g_1, g_2 \in G$, then their composition $f_1 \circ f_2 = f_1(f_2(\cdots))$ also satisfies strict equivariance. Since $f_1(g_1 \cdot x) = g_1 \cdot f_1(x)$ and $f_2(g_2 \cdot x) = g_2 \cdot f_2(x)$, we have $f_2(f_1(g_1 \cdot x)) = f_2(g_1 \cdot f_1(x)) = g_1 \cdot f_2(f_1(x))$, completing the proof. The challenge of a strict equivariant network is in designing equivariant layers. Two typical methods are raised by weight sharing (Cohen and Welling 2016) and weight tying (Cohen and Welling 2017).

Relaxed Equivariant Network. A strict equivariant network considers equivariance but assumes uniform and strict equivariance from the original data. However, real-world data rarely conforms to strict equivariance. To address this problem, we relax the G-transformation filters to realize relaxed equivariance. Also, like above, given a set of relaxed equivariant functions $\{\tilde{f}_i\}$, a relaxed equivariant network can be the composition function of $\{\tilde{f}_i\}$. The proof can be referred to (Kaba and Ravanbakhsh 2023).

Group Equivariant Convolution (GConv)

GConv achieves equivariant inductive biases by sharing weights convolution filters under group transformations. In a special case, CNNs achieve translation equivariance through translation transformations on the plane $\mathbb{Z}^2$.

To begin, we define the group operator $\varphi_G(\cdot)$ performs the G-transformation in the last two dimensions and cyclical permutations in the input channel dimension for $\cdot$, and the symbol $[\cdots]$ denotes the Pytorch style index operation. For convenience, we also define C_l, k_l, h_l, and w_l denote the channel number, filter size, width, and height of the 2D input or output in the l-layer, respectively. These definitions are used in the following context.

Lift Convolution. The first layer of G-CNNs typically lifts the input on the plane $\mathbb{Z}^2$ to the group G. Assume the input $\mathcal{Y}_1$ of size $[C_1, h_1, w_1]$ and the initial weight $\mathcal{W}_1$ with Kaiming Distribution of size $[C_2, C_1, k_1, k_1]$ on the plane $\mathbb{Z}^2$ in the first layer. Therefore, we obtain the full lift convolution filter $\mathcal{F}_1 = \varphi_G(\mathcal{W}_1)$ of size $[C_2, G_2, C_1, k_1, k_1]$ that contains an additional dimension G_2 for the output group. Note that $\mathcal{F}_1$ is constructed from $\mathcal{W}_1$ during each forward function. For all $u \in [1, C_2], v \in [1, G_2]$, the lift convolution can be performed by convolving over the input channel C_1 and summing up the outputs as follows:

$$\mathcal{Y}_2[u, v, :, :] = \sum_{m}^{C_1} \mathcal{Y}_1[m, :, :] * \mathcal{F}_1[u, v, m, :, :], \tag{3}$$

with the size of the output $\mathcal{Y}_2$ is $[C_2, G_2, h_2, w_2]$.

Group Convolution. Unlike the input on the plane $\mathbb{Z}^2$, GConv typically encodes the added group G in an extra tensor dimension. Assume the input $\mathcal{Y}_l$ of size $[C_l, G_l, h_l, w_l]$ on the group G, where G_l denotes the dimension of G in the l-layer ($l \geq 2$), and the initial weight $\mathcal{W}_l$ with Kaiming Distribution of size $[C_{l+1}, C_l, G_l, k_l, k_l]$ contains an additional dimension G_l for the input group. Then, we obtain the full group convolution filter $\mathcal{F}_l = \varphi_G(\mathcal{W}_l)$ of size $[C_{l+1}, G_{l+1}, C_l, G_l, k_l, k_l]$ containing an additional dimension G_{l+1} for the output group. For all $u \in [1, C_{l+1}], v \in [1, G_{l+1}]$, the group convolution can be performed by convolving over the input channel C_l and input group dimension G_l, and summing up the outputs as follows:

$$\mathcal{Y}_{l+1}[u, v, :, :] = \sum_{m}^{C_l} \sum_{n}^{G_l} \mathcal{Y}_l[m, n :, :] * \mathcal{F}_l[u, v, m, n, :, :], \tag{4}$$

with the size of the output $\mathcal{Y}_{l+1}$ is $[C_{l+1}, G_{l+1}, h_{l+1}, w_{l+1}]$. Since the group convolution is the function $f : G \to G$, we have $G_{l+1} = G_l = \text{Dim}(G)$, where the $\text{Dim}(G)$ denotes the dimension of G (e.g., 2 / 4 / 8 on the group $\mathcal{C}_2$ / $\mathcal{C}_4$ / $\mathcal{C}_8$).

Proposed Method

We focus on a specific case, i.e., Relaxed Rotational Equivariance (RRE) for Rotational Symmetry-Breaking (RSB) on the discrete rotational group $\mathcal{C}_n$. Then, we introduce the method of the proposed Relaxed Rotational Equivariant Conolution (RREConv) and the model for both classification and detection tasks based on RREConv in 2D vision.

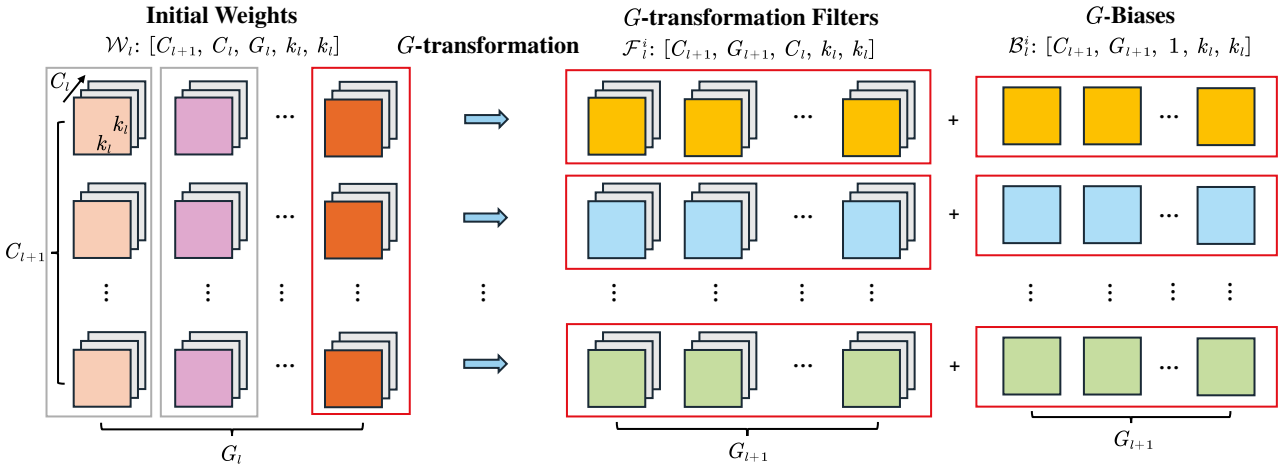


Figure 3: The construction of Relaxed Rotational Equivariant Filter (RREF). Note that the initial weights have G_l columns, but only the G-transformation in the last column (**Red Box**) is shown here for the convenience of drawing. The operations in other columns (**Gray Box**) are the same as the last column (**Red Box**).

Relaxed Rotational Equivariant Filter (RREF)

The construction of RREF is the key to our method. Assume an initial weight $\mathcal{W}_l$ of size $[C_{l+1}, C_l, G_l, k_l, k_l]$ with Kaiming Distribution on the group G in the l-layer, where $G = \mathcal{C}_n$ and $G_l = n$ especially. Define a set of affine matrices $\mathcal{A} = \{\mathcal{A}_i \mid i \in \{0, 1, \cdots, n-1\}\}$ for the G-transformation, where

$$\mathcal{A}_i = \begin{bmatrix} \cos(2\pi i/n) & -\sin(2\pi i/n) \\ \sin(2\pi i/n) & \cos(2\pi i/n) \end{bmatrix}. \tag{5}$$

Before G-transformation for coordinates, assume the coordinate system is at the centre of the 2D plane and define a function $\texttt{CoorSet}(\cdot)$ to obtain the set of all coordinates of $\cdot$. For all $u \in [1, C_{l+1}], m \in [1, C_l], n \in [1, G_l]$, 2D coordinate pair $(\mathbf{x}, \mathbf{y}) \in \texttt{CoorSet}(\mathcal{W}_l[u, m, n, :, :])$, we can obtain new 2D coordinate pair $(\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ after G-transformation for coordinates on $\mathcal{W}_l$ as follows:

$$(\tilde{\mathbf{x}_i}, \tilde{\mathbf{y}_i})^\top = \mathcal{A}_i\,(\mathbf{x}, \mathbf{y})^\top, \ \ \forall i \in \{0, 1, \cdots, n-1\} \tag{6}$$

Now, we obtain the i-filter of SRE on G as follows:

$$\mathcal{F}_l^i[u, m, n, \tilde{\mathbf{x}_i}, \tilde{\mathbf{y}_i}] = \mathcal{W}_l[u, m, n, \mathbf{x}, \mathbf{y}], \tag{7}$$

if $(\tilde{\mathbf{x}_i}, \tilde{\mathbf{y}_i}) \in \texttt{CoorSet}(\mathcal{F}_l[u, m, n, :, :])$. Note that some out-of-bounds coordinates of $\mathcal{F}_l^i$ may occur in some groups (e.g., $\mathcal{C}_6$ and $\mathcal{C}_8$) except $\mathcal{C}_2$ and $\mathcal{C}_4$. Based on this reason, some coordinates remain unassigned after the G-transformation from $\mathcal{W}_l$. For these coordinates, we employ Bilinear Interpolation. To achieve RREF, we introduce a set of learnable G-Biases $\mathcal{B}_l = \{\mathcal{B}_l^i \mid i \in \{0, 1, \cdots, n-1\}\}$, where the size of $\mathcal{B}_l^i$ is $[C_{l+1}, 1, 1, k_l, k_l]$ with Zero Distribution in the l-layer. Note that $\mathcal{B}_l$ can be updated end-to-end during the training period to adapt RSB in the dataset. Thus, the final values of $\mathcal{B}_l$ in l-layer are determined by datasets. Then, we obtain the i-filter of RRE on G as follows:

$$\mathcal{R}_l^i[u, m, n, :, :] = \mathcal{F}_l^i[u, m, n, :, :] + \mathcal{B}_l^i[u, 1, 1, :, :]. \tag{8}$$

Thus, the full RREF in the l-layer can be stacked as follows:

$$\mathcal{R}_l = Stack(\{\mathcal{R}_l^i \mid i \in \{0, 1, \cdots n-1\}\}), \tag{9}$$

with the size of $[C_{l+1}, G_{l+1}, C_l, G_l, k_l, k_l]$. An easily understandable construction of RREF can be seen in Figure 3.

Relaxed Rotational Equivariant Convolution

Consistent with the operator in Eq. (3) and Eq. (4), our Relaxed Rotational Lift Convolution (RRLConv) and Relaxed Rotational Equivariant Convolution (RREConv) can be written as Eq. (10) and Eq. (11) below, respectively:

$$\mathcal{Y}_2[u, v, :, :] = \sum_m^{C_1} \mathcal{Y}_1[m, :, :] * \mathcal{R}_1[u, v, m, :, :]. \tag{10}$$

$$\mathcal{Y}_{l+1}[u, v, :, :] = \sum_m^{C_l} \sum_n^{G_l} \mathcal{Y}_l[m, n\ :, :] * \mathcal{R}_l[u, v, m, n, :, :]. \tag{11}$$

Proof and Analysis

Conclusion 1. We say RREConv satisfies Eq. (2).

Proof 1. On the group $G = \mathcal{C}_n$, we define the operator $\varphi_G^j(\cdot)$ that rotates $\cdot$ by $2\pi i/n$. For any $j \in \{0, 1, \cdots, n-1\}$, we have two calculations as follows:

$$\begin{aligned} \mathcal{Y}_{l+1}(\varphi_G^j(\mathcal{Y}_l)) &= \sum_m^{C_l} \sum_n^{G_l} \varphi_G^j(\mathcal{Y}_l) * \mathcal{R}_l \\ &= \sum_m^{C_l} \sum_n^{G_l} \varphi_G^j(\mathcal{Y}_l) * \mathcal{F}_l + \sum_m^{C_l} \sum_n^{G_l} \varphi_G^j(\mathcal{Y}_l) * \mathcal{B}_l. \end{aligned} \tag{12}$$

$$\begin{aligned} \varphi_G^j(\mathcal{Y}_{l+1}(\mathcal{Y}_l)) &= \sum_m^{C_l} \sum_n^{G_l} \varphi_G^j(\mathcal{Y}_l) * \varphi_G^j(\mathcal{R}_l) \\ &= \sum_m^{C_l} \sum_n^{G_l} \varphi_G^j(\mathcal{Y}_l) * \varphi_G^j(\mathcal{F}_l) + \sum_m^{C_l} \sum_n^{G_l} \varphi_G^j(\mathcal{Y}_l) * \varphi_G^j(\mathcal{B}_l). \end{aligned} \tag{13}$$

Since $\mathcal{C}_n$ is a cyclic group, $\forall i \in \{0, 1, \cdots, n-1\}$, we have:

$$\varphi_G^j(\{\mathcal{F}_l^i\}) = \{\varphi_G^j(\mathcal{F}_l^i)\} = \{\mathcal{F}_l^{(i+j)\ \%\ n}\} = \{\mathcal{F}_l^i\},$$
$$\text{then, } \sum_m^{C_l}\sum_n^{G_l} \varphi_G^j(\mathcal{Y}_l) * \varphi_G^j(\mathcal{F}_l) = \sum_m^{C_l}\sum_n^{G_l} \varphi_G^j(\mathcal{Y}_l) * \mathcal{F}_l. \tag{14}$$

Then we have the L2-Norm as follows:

$$\|\mathcal{Y}_{l+1}(\varphi_G^j(\mathcal{Y}_l)) - \varphi_G^j(\mathcal{Y}_{l+1}(\mathcal{Y}_l))\| = \sum_m^{C_l}\sum_n^{G_l} \varphi_G^j(\mathcal{Y}_l) * (\mathcal{B}_l - \varphi_G^j(\mathcal{B}_l)) \leq \varepsilon. \tag{15}$$

Thus, the proposed RREConv satisfies Eq. (2). Especially when $\mathcal{B}_l = 0$, Eq. (15) equals 0 (i.e., $\varepsilon = 0$), where RREConv satisfies Eq. (1). Note that $\mathcal{B}_l$ updates end-to-end. Only at the beginning of the training period, $\mathcal{B}_l = 0$.

Projection Error. Considering Eq. (11) of RREConv and Eq. (4) of GConv, we can obtain the Projection Error (PE):

$$\|\sum_m^{C_l}\sum_n^{G_l}(\mathcal{Y}_l * \mathcal{R}_l - \mathcal{Y}_l * \mathcal{F}_l)\| = \|\sum_m^{C_l}\sum_n^{G_l}\mathcal{Y}_l * \mathcal{B}_l\|. \tag{16}$$

Therefore, each RREConv in the RRE network aims to optimize PE to adopt RSB in natural image datasets.

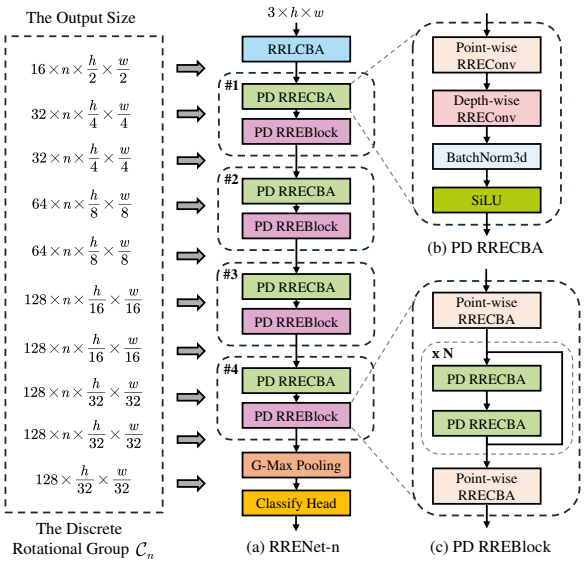


Figure 4: The architecture of the backbone RRENet-n based on RREConv. Note that n denotes the dimension of $\mathcal{C}_n$, and "-CBA" means "Conv + BatchNorm + Activate" operations.

Model Architecture

Relaxed Rotational Equivariance Network (RRENet). We propose the RRENet based on RREConv. Considering significant computational and parameter overhead caused by the additional G-dimension, we redesign the Point-wise and Depth-wise versions of RREConv. We adopt the classic structure of ResNet (He et al. 2016), where each PD RREBlock consists of two Point-wise RRECBAs adjusting the number of channels, with two PD RRECBAs in between, and residual connections are used, as shown in Figure 4.

Relaxed Rotational Equivariance Detector (RREDet). We also propose the RREDet, where we use the RRENet as the backbone to obtain the initial three scale features in {2, 3, 4}-layer, and the FPN+PAN architecture as the neck layer to obtain the final three scale features with size 80×80, 40×40, and 20×20 for different-size object detection. Note that the feature maps in 4-layer are input to the G-SPPF for Spatial Pyramid Pooling (He et al. 2015) on the group $\mathcal{C}_n$. Consequently, these features are fed into the G-Max Pooling and the Detector Head, as shown in Figure 5.

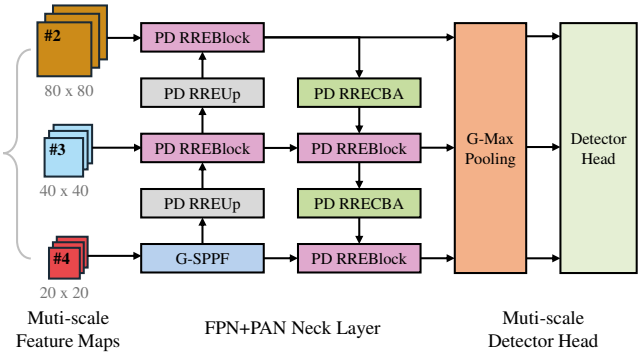


Figure 5: The architecture of RREDet. Note that #{2,3,4} denote muti-scale feature maps from {2,3,4}-layer in the backbone RRENet. The PD RREUp adopts the same structure as the PD RREConv, except for transposed convolution.

Experiments

In this section, we conduct extensive ablation experiments to demonstrate the effectiveness of the proposed RREConv in both classification and 2D object detection tasks. All the parameters are set the same, and all the experiments are conducted on the dual RTX-4090 GPUs.

We evaluate the proposed method on the CIFAR10 / 100, PASCAL VOC07+12, and MS COCO2017 datasets. The ablation experiments mentioned above on both tasks show that the proposed method for RRE with fewer additional parameters achieves reasonable performance growth compared to the method for SRE. Also, the proposed method does not increase training and inference time.

Training Details

All training in this paper is based on the famous engine library "Ultralytics" (Jocher, Qiu, and Chaurasia 2023).

Data Augmentation. Like Ultralytics, we employ the default data augmentation settings such as Erasing of 0.4, Mosaic of 1.0, HSV-Saturation of 0.7, HSV-Value of 0.4, and HSV-Hue of 0.015 for the 2D object detection tasks. We also use random augmentation for classification tasks.

Group G	Type of Equivariance	CIFAR-10		CIFAR-100		PASCAL VOC07+12		
		Top-1 Acc.	#Param.	Top-1 Acc.	#Param.	AP_{50}	$AP_{50:95}$	#Param.
$\mathbb{Z}^2$	NRE	95.6	1.08M	77.2	1.19M	79.1	58.6	3.11M
$\mathcal{C}_2$	SRE	94.4_{base}	0.49M	76.1_{base}	0.61M	78.9_{base}	58.3_{base}	1.81M
	RRE	$\mathbf{95.0}_{+0.6}$	0.51M	$\mathbf{76.6}_{+0.5}$	0.63M	$\mathbf{80.1}_{+1.2}$	$\mathbf{59.7}_{+1.4}$	1.85M
$\mathcal{C}_4$	SRE	95.6_{base}	0.79M	80.1_{base}	0.91M	83.1_{base}	64.1_{base}	2.83M
	RRE	$\mathbf{96.5}_{+0.9}$	0.83M	$\mathbf{80.9}_{+0.8}$	0.95M	$\mathbf{84.1}_{+1.0}$	$\mathbf{65.2}_{+1.1}$	2.91M
$\mathcal{C}_6$	SRE	96.0_{base}	1.09M	80.6_{base}	1.21M	83.8_{base}	64.7_{base}	3.85M
	RRE	$\mathbf{96.8}_{+0.8}$	1.16M	$\mathbf{81.3}_{+0.7}$	1.27M	$\mathbf{84.5}_{+0.7}$	$\mathbf{65.5}_{+0.8}$	3.98M
$\mathcal{C}_8$	SRE	96.5_{base}	1.39M	82.1_{base}	1.39M	85.2_{base}	66.6_{base}	4.88M
	RRE	$\mathbf{97.2}_{+0.7}$	1.48M	$\mathbf{82.7}_{+0.6}$	1.48M	$\mathbf{86.0}_{+0.8}$	$\mathbf{67.5}_{+0.9}$	5.04M

Table 1: Ablation experiments compared with NRE, SRE, and RRE on the group $\mathcal{C}_n (n = 2, 4, 6, 8)$ on the CIFAR10 / 100 datasets and the PASCAL VOC07+12 dataset. All models are trained from scratch on the architecture of the proposed RRENet for classification tasks and the proposed RREDet for 2D object detection tasks.

Training Settings. All models are trained from scratch for 300 epochs for the 2D object detection tasks and 100 epochs for the classification tasks. We also default to using an SGD optimizer for both tasks with an initial learning rate of 0.01, a final learning rate of 0.01, a momentum of 0.937, a weight decay of 5e-4, a warmup epoch of 3, a warmup momentum of 0.8, and a warmup bias learning rate of 0.1.

Experimental Results

Ablation Experiments in Classification. To evaluate the effectiveness of the proposed RREConv, we conduct extensive ablation experiments on the CIFAR10 / 100 datasets, using the architecture of RRENet with Non-Rotational Equivariance (NRE) on the plane $\mathbb{Z}^2$, with SRE and RRE on the group $\mathcal{C}_n$. For SRE, we replace all RREConv with GConv. For NRE, we replace all RREConvs with vanilla Convs but remove G-Max Pooling. We keep all training parameters consistent in NRE, SRE, and RRE.

The results can be seen in Table 1. From the table, the top-1 accuracy of the model with RRE consistently outperforms their NRE and SRE counterparts on the $\mathcal{C}_4$, $\mathcal{C}_6$, and $\mathcal{C}_8$ groups. On the group $\mathcal{C}_2$, although the top-1 accuracy of the models with RRE and SRE is lower than that with NRE, their parameters are only half of the model with NRE. The results demonstrate that RRE achieves better results in classification tasks while maintaining a minor parameter increase compared to SRE. In addition, we find that the model with SRE or RRE on the group $\mathcal{C}_4$ achieves the trade-off between parameters and performance.

Ablation Experiments in 2D Object Detection. We also conduct extensive ablation experiments on the PASCAL VOC07+12 dataset to validate the effectiveness of the proposed RREConv in 2D object detection tasks using the standard Average Precision (AP) metric. We typically employ the AP_{50} metric at an Intersection over Union (IoU) threshold of 0.5, and the $AP_{50:95}$ metric across IoU thresholds ranging from 0.5 to 0.95 as key evaluation metrics.

As shown in Table 1, the AP_{50} and $AP_{50:95}$ of the model with RRE consistently surpasses their NRE and SRE counterparts on all groups. The results prove that the proposed method for RRE can also achieve better results in 2D object detection tasks and maintain less parameter increase compared to the method for RRE. Like the results in classification tasks, the models with SRE or RRE on the group $\mathcal{C}_4$ also balance parameters and performance.

Model	Top-1 Acc.	#Param.
WideResNet	79.5	36.5M
ResNeXt-29	82.7	68.1M
DenseNet-BC	82.8	25.6M
Res2NeXt-29	83.2	36.7M
RRENet-n ($\mathcal{C}_4$)	80.9	0.95M
RRENet-s ($\mathcal{C}_4$)	83.5	2.97M
RRENet-m ($\mathcal{C}_4$)	**85.0**	8.66M

Table 2: Top-1 Accuracy (%) on the CIFAR-100 dataset. All models are trained from scratch.

Comparison with other models in Classification. Table 2 presents comparative results of three different-size RRENet models (i.e., RRENet-n / m / s) on the group $\mathcal{C}_4$ against other classic convolutional models, including WideResNet, ResNeXt-29, DenseNet-BC, and Res2NeXt-29, on the CIFAR-100 dataset. From the table, RRENet-m ($\mathcal{C}_4$) achieves a significant improvement in the top-1 accuracy, with a range of 2.2%-6.9% enhancement over others, but its parameters are only 13%-34% of others. RRENet-s ($\mathcal{C}_4$) still surpasses others, but it has fewer parameters, only one-third that of RRENet-m ($\mathcal{C}_4$), achieving a better balance between parameters and accuracy. Although RRENet-n ($\mathcal{C}_4$) only exceeds WideResNet over others, its parameters are only 0.95M, which is only 2.6% that of WideResNet.

Comparison with other models in 2D Object Detection. As shown in Table 3, three different-size RRENet models (i.e., RREDet-n / m / s) on the group $\mathcal{C}_4$ mainly compares

Model	AP_{50}	$AP_{50:95}$	#Param.
YOLOv8-n	78.6	57.5	3.0M
YOLOv8-s	81.6	61.6	11.1M
YOLOv8-m	83.7	65.3	25.9M
RREDet-n ($\mathcal{C}_4$)	84.1	65.2	2.9M
RREDet-s ($\mathcal{C}_4$)	86.3	67.6	10.5M
RREDet-m ($\mathcal{C}_4$)	**87.4**	**70.3**	25.7M

Table 3: Average Precise (AP) on the PASCAL VOC07+12 dataset. All models are trained from scratch.

Model	AP_{50}	$AP_{50:95}$	#Param.
YOLOv5-s	**56.8**	37.4	7.2M
YOLOv6-n	53.1	37.5	4.7M
YOLOv7-tiny	55.2	37.4	6.2M
YOLOv8-n	52.6	37.3	3.2M
YOLOv9-n	53.1	38.3	2.0M
YOLOv10-n	-	38.5	2.3M
RREDet-n ($\mathcal{C}_4$)	55.2	**40.2**	3.1M

Table 4: Average Precise (AP) on the MS COCO2017 dataset. All models are trained from scratch.

with the advanced YOLOv8 with three size models (i.e., YOLOv8-n / m / s) on the PASCAL VOC07+12 dataset. Among them, RREDet-n ($\mathcal{C}_4$) achieves approximate AP_{50} and $AP_{50:95}$ compared to YOLOv8-m, but its parameters are only 11% of YOLOv8-m. In addition, RREDet-s / m ($\mathcal{C}_4$) exceeds other models in both AP_{50} and $AP_{50:95}$. Although RREDet-m ($\mathcal{C}_4$) has slightly higher AP metric than RREDet-s ($\mathcal{C}_4$), its parameters are $2.45\times$ that of RREDet-m ($\mathcal{C}_4$). RREDet-s ($\mathcal{C}_4$) balances AP and parameters.

Furthermore, we conduct experiments on the larger-scale MS COCO2017 dataset to test the proposed RREDet models' generalization ability. We compare RREDet-n ($\mathcal{C}_4$) with other YOLO family models of the same scale, as shown in Table 4. From the table, RREDet-n ($\mathcal{C}_4$) outperforms the other models mentioned above in $AP_{50:95}$, but is lower than YOLOv5-s in AP_{50}. Nevertheless, the parameters of YOLOv5-s are $2\times$ that of RREDet-n ($\mathcal{C}_4$). The latest models YOLOv9-n / YOLOv10-n have approximately 65% / 74% parameters of RREDet-n ($\mathcal{C}_4$), but RREDet-n ($\mathcal{C}_4$) improve around 5% / 4.4% in $AP_{50:95}$.

Summary. The experiments above prove the advancement of the proposed method for RRE in 2D vision. The RSB is common in the real world. Therefore, relaxing SRE to obtain RRE is an effective way to adapt to RSB. Since the proposed G-Biases are learnable parameters that can be updated end-to-end during the training period, they are automatically updated based on the distribution characteristics of the natural dataset to adapt to RSB. In Pytorch, we use *torch.nn.Parameter*$(\cdots,$ *requires_grad=True*$)$ to define G-Biases that can be updated by gradient descent direction. Overall, RRE networks can achieve better results than traditional SRE networks in 2D vision.

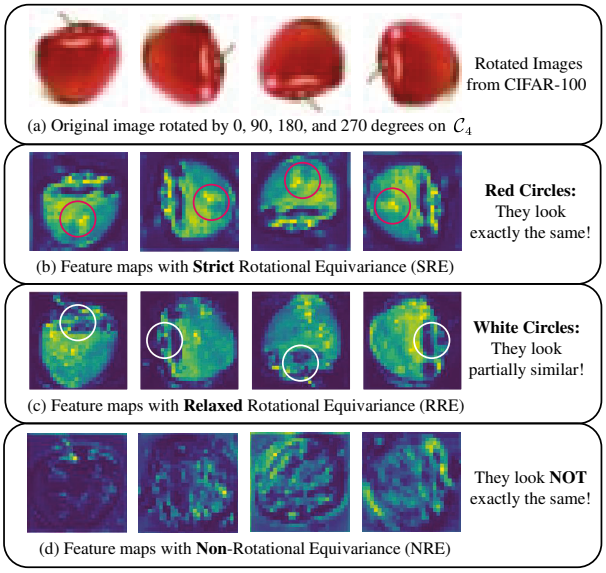


Figure 6: The visualization of SRE, RRE, and NRE.

Visualization of SRE, RRE, and NRE.

The visualization of SRE, RRE, and RRE on the group $\mathcal{C}_4$ can be seen in Figure 6. In (b), we observe that the content inside red circles remained unchanged after rotation, and the overall image shows the same, reflecting SRE's characteristics. In (c), we find slight differences in the content inside white circles after rotation, but the overall image presents similarity, reflecting RRE's characteristics. In (d), they are almost different in the content. The similarity of RRE, rather than uniformity, is more common and primarily reflects the RSB of objects in the real world.

Conclusion

This paper delves into rotational equivariant networks for modelling natural datasets, highlighting their effectiveness in leveraging rotation groups. Despite their advancements, existing networks capture uniform and strict rotational symmetry from original data, which does not align with real-world data characterized by Rotational Symmetry-Breaking (RSB). This inability to effectively adapt RSB scenarios within natural datasets necessitates a novel method.

To tackle this challenge, we introduce a simple yet powerful method involving a set of learnable parameters called G-Biases. Utilizing this innovative mechanism, we propose a Relaxed Rotational Equivariant Convolution (RREConv), tailored to address the nuances of RSB. The extensive experiments have proven the effectiveness of our method.

Exploring Symmetry-Breaking with relaxed equivariance within the realms of 2D and 3D visual fields represents a compelling avenue for future research. We firmly believe incorporating this relaxation principle into strict equivariant models can improve their ability to represent Symmetry-Breaking phenomena in real-world contexts.

Acknowledgments

This research is supported by the National Natural Science Foundation of China (No.42130112, No.42371479), General Program of Shanghai Natural Science Foundation(Grant No.24ZR1419800, No.23ZR1419300), Science and Technology Commission of Shanghai Municipality (Grant No.22DZ2229004), Beijing Natural Science Foundation (No.QY23187), and Shanghai Frontiers Science Center of Molecule Intelligent Syntheses.

References

Barone, M.; and Theophilou, A. 2008. Symmetry and symmetry breaking in modern physics. In *Journal of Physics: Conference Series*, volume 104, 012037. IOP Publishing.

Bogatskiy, A.; Ganguly, S.; Kipf, T.; Kondor, R.; Miller, D. W.; Murnane, D.; Offermann, J. T.; Pettee, M.; Shanahan, P.; Shimmin, C.; et al. 2022. Symmetry group equivariant architectures for physics. *arXiv preprint arXiv:2203.06153*.

Chidester, B.; Do, M. N.; and Ma, J. 2018. Rotation equivariance and invariance in convolutional neural networks. *arXiv preprint arXiv:1805.12301*.

Cohen, T.; and Welling, M. 2016. Group equivariant convolutional networks. In *International conference on machine learning*, 2990–2999. PMLR.

Cohen, T. S.; and Welling, M. 2017. Steerable CNNs. In *International Conference on Learning Representations*.

Desai, K.; Nachman, B.; and Thaler, J. 2022. Symmetry discovery with deep learning. *Physical Review D*, 105(9): 096031.

Dieleman, S.; De Fauw, J.; and Kavukcuoglu, K. 2016. Exploiting cyclic symmetry in convolutional neural networks. In *International conference on machine learning*, 1889–1898. PMLR.

Esteves, C. 2020. Theoretical aspects of group equivariant neural networks. *arXiv preprint arXiv:2004.05154*.

Fuchs, F.; Worrall, D.; Fischer, V.; and Welling, M. 2020. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in neural information processing systems*, 33: 1970–1981.

Ghosh, S. K.; Biswas, P. K.; Xu, C.; Li, B.; Zhao, J. Z.; Hillier, A. D.; and Xu, X. 2022. Time-reversal symmetry breaking superconductivity in three-dimensional Dirac semimetallic silicides. *Physical Review Research*, 4(1).

Han, J.; Ding, J.; Xue, N.; and Xia, G.-S. 2021. Redet: A rotation-equivariant detector for aerial object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2786–2795.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 37(9): 1904–1916.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

He, L.; Chen, Y.; Dong, Y.; Wang, Y.; Lin, Z.; et al. 2021. Efficient equivariant network. *Advances in Neural Information Processing Systems*, 34: 5290–5302.

Jocher, G.; Qiu, J.; and Chaurasia, A. 2023. Ultralytics YOLO.

Kaba, S.-O.; Mondal, A. K.; Zhang, Y.; Bengio, Y.; and Ravanbakhsh, S. 2023. Equivariance with learned canonicalization functions. In *International Conference on Machine Learning*, 15546–15566. PMLR.

Kaba, S.-O.; and Ravanbakhsh, S. 2023. Symmetry Breaking and Equivariant Neural Networks. In *NeurIPS 2023 Workshop on Symmetry and Geometry in Neural Representations*.

Kavukcuoglu, K.; Ranzato, M.; Fergus, R.; and LeCun, Y. 2009. Learning invariant features through topographic filter maps. In *2009 ieee conference on computer vision and pattern recognition*, 1605–1612. IEEE.

Kondor, R.; and Trivedi, S. 2018. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *International conference on machine learning*, 2747–2755. PMLR.

Li, J.; Yang, Z.; Liu, H.; and Cai, D. 2018. Deep rotation equivariant network. *Neurocomputing*, 290: 26–33.

Li, Z.; Nagano, L.; and Terashi, K. 2024. Enforcing exact permutation and rotational symmetries in the application of quantum neural networks on point cloud datasets. *Physical Review Research*, 6(4): 043028.

Liu, H.; Yang, J.; Li, Z.; Li, K.; Zheng, J.; Wang, X.; Tang, X.; Chen, M.; You, X.; and Wei, X. 2024. ProEqBEV: Product Group Equivariant BEV Network for 3D Object Detection in Road Scenes of Autonomous Driving. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, 16178–16184. IEEE.

Marcos, D.; Volpi, M.; Komodakis, N.; and Tuia, D. 2017. Rotation Equivariant Vector Field Networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE.

Marcos, D.; Volpi, M.; and Tuia, D. 2016. Learning rotation invariant convolutional filters for texture classification. In *2016 23rd International Conference on Pattern Recognition (ICPR)*, 2012–2017. IEEE.

Müller, P.; Golkov, V.; Tomassini, V.; and Cremers, D. 2021. Rotation-equivariant deep learning for diffusion MRI. *arXiv preprint arXiv:2102.06942*.

Ravanbakhsh, S.; Schneider, J.; and Poczos, B. 2017. Equivariance through parameter-sharing. In *International conference on machine learning*, 2892–2901. PMLR.

Romero, D. W.; and Lohit, S. 2022. Learning partial equivariances from data. *Advances in Neural Information Processing Systems*, 35: 36466–36478.

Seo, A.; Kim, B.; Kwak, S.; and Cho, M. 2022. Reflection and rotation symmetry detection via equivariant learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 9539–9548.

van der Ouderaa, T.; Romero, D. W.; and van der Wilk, M. 2022. Relaxing equivariance constraints with non-stationary

continuous filters. *Advances in Neural Information Processing Systems*, 35: 33818–33830.

Veeling, B. S.; Linmans, J.; Winkens, J.; Cohen, T.; and Welling, M. 2018. Rotation equivariant CNNs for digital pathology. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II 11*, 210–218. Springer.

Vernizzi, G.; and Wheater, J. F. 2002. Rotational symmetry breaking in multimatrix models. *Physical Review D*, 66(8): 085024.

Wang, R.; Hofgard, E.; Gao, H.; Walters, R.; and Smidt, T. 2024. Discovering Symmetry Breaking in Physical Systems with Relaxed Group Convolution. In *Forty-first International Conference on Machine Learning*.

Wang, R.; Walters, R.; and Yu, R. 2022. Approximately equivariant networks for imperfectly symmetric dynamics. In *International Conference on Machine Learning*, 23078–23091. PMLR.

Wang, X.; Lei, J.; Lan, H.; Al-Jawari, A.; and Wei, X. 2023. DuEqNet: dual-equivariance network in outdoor 3D object detection for autonomous driving. In *2023 IEEE International conference on robotics and automation (ICRA)*, 6951–6957. IEEE.

Weiler, M.; and Cesa, G. 2019. General e (2)-equivariant steerable cnns. *Advances in neural information processing systems*, 32.

Weiler, M.; Geiger, M.; Welling, M.; Boomsma, W.; and Cohen, T. S. 2018. 3d steerable cnns: Learning rotationally equivariant features in volumetric data. *Advances in Neural Information Processing Systems*, 31.

Wu, F.; Hu, P.; and Kong, D. 2015. Flip-rotate-pooling convolution and split dropout on convolution neural networks for image classification. *arXiv preprint arXiv:1507.08754*.

Zhu, K.; Zhang, X.; Chen, G.; Li, X.; Cai, P.; Liao, P.; and Wang, T. 2022. Multi-oriented rotation-equivariant network for object detection on remote sensing images. *IEEE Geoscience and Remote Sensing Letters*, 19: 1–5.