

# “Rotation Equivariant Arbitrary Scale Image Super-Resolution”: Supplementary Material

Qi Xie, Jiahong Fu, Zongben Xu and Deyu Meng

## Abstract

This supplementary material presents the rigorous definition on image rotation with different network layers and the proofs of all involved theoretical results in the main paper. More intuitive illustration of the theoretical results, implementation details of the proposed method and more experimental results are also provided for more comprehensive reference of readers.

## 1 RIGOROUS DEFINITION ON IMAGE ROTATION

In order to rigorously define image rotation, an  $h \times w \times c$  image  $I$  need to be viewed as a two-dimensional discretization of a smooth function  $r : \mathbb{R}^2 \rightarrow \mathbb{R}^k$ , at the cell-center of a regular grid with  $h \times w \times c$  cells [1], i.e., for  $\forall i = 1, 2, \dots, h$ ,  $\forall j = 1, 2, \dots, w$ , and  $\forall c = 1, 2, \dots, n_0$

$$I_{ij}^c = r_c(x_{ij}), \quad (1)$$

where  $x_{ij} = ((i - \frac{h+1}{2}) \delta, (j - \frac{w+1}{2}) \delta)^T$ ,  $r(x) = [r_1(y), r_2(x), \dots, r_c(x)]$  and  $\delta$  is the mesh size of the image grid. Similar to  $I$ , we represent the feature maps of equivariant convolution as  $F \in \mathbb{R}^{h \times w \times n \times t}$ , where  $t = |S|$  and  $S$  is the equivariant group. Specifically,  $F$  is a four-dimensional grid function, whose  $k^{th}$  channel is sampled from a smooth function  $e_c : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$ , i.e., for  $i = 1, 2, \dots, h$ ,  $j = 1, 2, \dots, w$  and  $c = 1, 2, \dots, n$

$$F_{ij}^{c,A} = e_c(x_{ij}, A), \quad (2)$$

where  $x_{ij} = ((i - \frac{h+1}{2}) \delta, (j - \frac{w+1}{2}) \delta)^T$ ,  $e(x, A) = [e_1(x, A), e_2(x, A), \dots, e_n(x, A)]$  and  $A \in S$ .

Then, in continuous domain, the rotation is easy to formally define. Specifically, for an input  $r \in C^\infty(\mathbb{R}^2)$  and a rotation degree  $\theta \in [0, 2\pi]$ , let  $A_\theta \in O(2)$  be the rotation matrix  $\begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ .  $A$  acts on  $r$  by

$$\tilde{\pi}_{\tilde{A}}^I[r](x) = r(A_\theta^{-1}x), \quad \forall x \in \mathbb{R}^2. \quad (3)$$

For a feature map  $e \in C^\infty(E(2))$ ,  $E(2) = \mathbb{R}^2 \ltimes O(2)$ , and a degree  $\theta \in [0, 2\pi]$ .  $A_\theta$  acts on  $e$  by

$$\tilde{\pi}_{\tilde{A}}^F[e](x, A) = e(A_\theta^{-1}x, A_\theta^{-1}A), \quad \forall (x, A) \in E(2). \quad (4)$$

Based on this we can further define the rotation on discrete domain. Specifically, the transformations  $\tilde{A} \in S$  (rotation for  $2\pi/t, 4\pi/t, \dots, 2t\pi/t$  degree) on each channel of the input image and the feature map are defined by

$$(\pi_{\tilde{A}}^I(I))_{ij}^c = \tilde{\pi}_{\tilde{A}}^I[r_c](x_{ij}), \quad (\pi_{\tilde{A}}^F(F))_{ij}^{c,A} = \tilde{\pi}_{\tilde{A}}^F[e_c](x_{ij}, A), \quad \forall i = 1, 2, \dots, h, j = 1, 2, \dots, w, c = 1, 2, \dots, n, A \in S, \quad (5)$$

where  $[\cdot]$  denotes the composition of functions. It is easy to find that, the rotation on feature map is consistent of spatially rotation on the first 2 dimensions and cyclically shifting along the final dimensions. Formally, for  $\forall \tilde{A} \in S$ , the rotation of feature map can be rewrite as

$$\pi_{\tilde{A}}^F(F) = [\pi_{\tilde{A}}^I(F^{\tilde{A}^{-1}A_1}), \pi_{\tilde{A}}^I(F^{\tilde{A}^{-1}A_2}), \dots, \pi_{\tilde{A}}^I(F^{\tilde{A}^{-1}A_t})], \quad (6)$$

where  $F^{A_k}$ ,  $k = 1, 2, \dots, t$  are tensors of size  $h \times w \times n$ , which is viewed as a  $n$ -channel image when performing spatially rotation  $\pi_{\tilde{A}}^I$ . Moreover, by substituting  $F_{ij}$  into (6), we have

$$\pi_{\tilde{A}}^F(F_{ij}) = [F_{ij}^{\tilde{A}^{-1}A_1}, F_{ij}^{\tilde{A}^{-1}A_2}, \dots, F_{ij}^{\tilde{A}^{-1}A_t}]. \quad (7)$$

For the proposed modules, the rotations on feature functions  $H_{ij}(x, A)$  and output function  $f_{ij}(x)$  are defined as following: for  $\forall x \in \mathbb{R}^2$  and  $A \in S$ ,

$$\begin{aligned} \pi_{\tilde{A}}^H[H_{ij}](x, A) &= H_{ij}(\tilde{A}^{-1}x, \tilde{A}^{-1}A), \\ \pi_{\tilde{A}}^f[f_{ij}](x) &= f_{ij}(\tilde{A}^{-1}x), \end{aligned} \quad (8)$$

where  $[\cdot]$  denotes the composition of functions and  $\tilde{A} \in S$ .

---

Qi Xie, Jiahong Fu, Zongben Xu and Deyu Meng (corresponding author) are with School of Mathematics and Statistics and Ministry of Education Key Lab of Intelligent Networks and Network Security, I'an Jiaotong University, Shaanxi, P.R.China.  
Email: xie.qi@mail.xjtu.edu.cn, jiahongfu@stu.xjtu.edu.cn, zbxu@mail.xjtu.edu.cn, eud.cn, dymeng@mail.xjtu.edu.cn.

## 2 IMPLEMENTATION DETAILS OF THE PROPOSED FRAMEWORK IN APPLICATION

**Input layer for Rot-E INR.** In the main text, we denote the operation of the input layer of Rot-E INR as  $H_{ij}(x, B) = [W_{\text{in}} \hat{\phi}_\varphi F_{ij}](x, B)$ , which is calculated by

$$[W_{\text{in}} \hat{\phi}_\varphi F_{ij}](x, B) = \sum_{A \in S} \varphi \left( W_{\text{in}}^{B^{-1}A}, F_{ij}^A, A^{-1}x \right), \quad (9)$$

where  $\varphi$  can be arbitrary function chosen for constructing INR and  $W_{\text{in}} = \{W_{\text{in}}^A | A \in S\}$  denotes the possible learnable parameters in this layer, which can be divided into  $t$  groups in similar structure, corresponding to the elements in  $S$ . As shown in Fig. 5(d) of the main text, when implementing the input layer,  $W_{\text{in}}$ , its elements are shifted cyclically along group index  $A$ , just as the convolution filters are shifted cyclically along the group index in the G-CNN framework [2].

It should be noted that the above process only calculates the INR for a local region. In practice, we need to perform the above calculation for each pixel of the high-resolution image to obtain the complete high-resolution image. Therefore, for the coordinates of each pixel in the high-resolution image, we need to sample its features on the low-resolution image grid and calculate the relative coordinates for performing (13). In general, we can implement the input layer of the INR through the following algorithm:

---

### Algorithm 1 Implementation of the input layer of Rot-E INR

---

**Initialize learnable parameters:** Initialize  $W \in \mathbb{R}^{m \times n \times t}$ , where  $m, n$  and  $t$  are the output channel, number input channel number, and transformation group size, respectively;

**Input:** Feature map  $F \in \mathbb{R}^{b \times n \times t \times H_l \times W_l}$  and relative tensor  $X \in \mathbb{R}^{b \times 2 \times H \times W}$ , where  $b, H_l, W_l, H$  and  $W$  are the batch size, low-resolution-image height, low-resolution-image weight, high-resolution-image height and width, respectively;

- 1: For each coordinate in  $X$ , sample the features of its nearest neighbor in the low-resolution image grid to obtain  $\hat{F} \in \mathbb{R}^{b \times n \times t \times H \times W}$  and calculate the relative coordinate  $X_{\text{rel}} \in \mathbb{R}^{b \times 2 \times H \times W}$ ;
- 2: Concatenate all the inverse transformation matrix  $A^{-1}, A \in S$  together and get  $A_{\text{inv}} \in \mathbb{R}^{2 \times 2 \times t}$ ;
- 3: Perform coordinate rotation:  $\hat{X} = \text{einsum}('bdhw, edt \rightarrow bethw', X_{\text{rel}}, A_{\text{inv}})$ , then  $\hat{X} \in \mathbb{R}^{b \times 2 \times t \times H \times W}$ ;
- 4: **for**  $k = 1 : t$  **do**
- 5:   Perform parameter cyclically shifting:  $\hat{W} = \text{roll}(W, \text{shifts} = k)$ ;
- 6:   Perform main operation:  $H[:, :, k, :, :] = \sum_{i=1}^t \varphi \left( \hat{W}_i, \hat{F}[:, :, i, :, :], \hat{X}[:, :, i, :, :] \right)$ ;
- 7: **end for**

**Output:**  $H \in \mathbb{R}^{b \times m \times t \times H \times W}$  where  $m$  is output channel number.

---

In the algorithm, “einsum” and “roll” are the Einstein summation operation and rolling operation, respectively. Here, we follow PyTorch’s programming conventions to present the pseudocode for readers’ convenience.

**Intermediate layer for Rot-E INR.** The intermediate layer of Rot-E INR maps a group of functions  $H_{ij}(x, B)$  to another group of functions  $\hat{H}_{ij}(x, A)$ . Specifically, for  $\forall x, A, B, H_{ij}(x, B) \in \mathbb{R}^m$  and  $\hat{H}_{ij}(x, A) \in \mathbb{R}^n$ , where  $m$  and  $n$  are the channel numbers of  $H(x, B)$  and  $\hat{H}(x, A)$ , respectively. We denote the operation of the intermediate layer as  $\hat{H}_{ij}(x, A) = [W \circ H_{ij}](x, A)$ , which is calculated by

$$[W \circ H_{ij}](x, A) = \sum_{B \in S} W^{A^{-1}B} \cdot H_{ij}(x, B), \quad (10)$$

where  $W$  is the learnable parameters, consistent of  $W^A \in \mathbb{R}^{n \times m}$ ,  $A \in S$ , and  $\cdot$  represents the matrix multiplication. It should be noted that  $W$  is shifted cyclically along group index just as the calculation in (9).

In application, the operation (10) actually can be easily performed with the following concise algorithm:

---

### Algorithm 2 Implementation of the intermediate layer of Rot-E INR

---

**Initialize learnable parameters:** Initialize  $W \in \mathbb{R}^{n \times m \times t}$ , where  $n, m$  and  $t$  are the output channel, number input channel number, and transformation group size, respectively;

**Input:** Feature map  $H \in \mathbb{R}^{b \times m \times t \times H \times W}$ , where  $b, H$  and  $W$  are the batch size, high-resolution-image height and width, respectively;

- 1: Stack the cyclically shifted  $W$ :  $\hat{W} = \text{stack}([\text{roll}(W, \text{shifts} = k), \text{for } k = 1 : t], \text{dim} = 4)$ ;
- 2: Perform linear operation:  $\hat{H} = \text{einsum}('bmthw, nmst \rightarrow bnshw', H, \hat{W})$ ;

**Output:**  $\hat{H} \in \mathbb{R}^{b \times n \times t \times H \times W}$ .

---

In the algorithm, “stack” is the stacking operation, where we follow PyTorch’s programming conventions to present the pseudocode.

**Output layer for Rot-E INR.** The output layer of Rot-E INR maps a group of functions  $\hat{H}_{ij}(x, A)$  to 2D functions  $f_{ij}(x)$ . Specifically, we denote it as  $f_{ij}(x) = [W_{\text{out}} \circ \hat{H}_{ij}](x)$ , which is calculated by

$$[W_{\text{out}} \circ \hat{H}_{ij}](x) = \psi_{W_{\text{out}}^2} \left( \sum_{A \in S} W_{\text{out}}^1 \cdot \hat{H}_{ij}(x, A) \right), \quad (11)$$

where  $W_{\text{out}} = \{W_{\text{out}}^1, W_{\text{out}}^2\}$ , and  $W_{\text{out}} \in \mathbb{R}^{m \times n}$  is a learnable matrix with  $m$  being a selected feature channel number,  $\psi_{W_{\text{out}}^2}$  denotes a arbitrary function with  $W_{\text{out}}^2$  being its learnable parameters.

In application, we can perform the operation (11) with the following concise algorithm:

---

**Algorithm 3** Implementation of the output layer of Rot-E INR

---

**Initialize learnable parameters:** Initialize  $W \in \mathbb{R}^{m \times n \times \dots}$ , where  $m$  and  $n$  are the output channel, number input channel number, respectively;

**Input:** Feature map  $H \in \mathbb{R}^{b \times n \times t \times H \times W}$ , where  $b, t, H$  and  $W$  are the batch size, transformation group size, high-resolution-image height and width, respectively;

1: Perform linear operation:  $\hat{I} = \sum_{k=1}^t \text{einsum}('bnhw, mn \rightarrow bmnw', H[:, :, k, :, :], \hat{W})$ ;

2: Perform nonlinear operation:  $\hat{I} = \psi(\hat{I})$ ,  $\psi$  can be arbitrary network or opeartion;

**Output:** High-resolution image  $\hat{I}$ .

---

For more details, please refer to our code at <https://github.com/XieQi2015/Equivariant-ASISR>.

### 3 ILLUSTRATION OF THE KEY CONCEPTS AND THEORETICAL RESULTS

In the main text, we have presented multiple theoretical conclusions and conducted equivariance analysis on each module of the proposed network framework. To make the theoretical results more accessible to a wider audience, we offer an intuitive illustration of the key concepts and the theoretical results in Fig. 1.

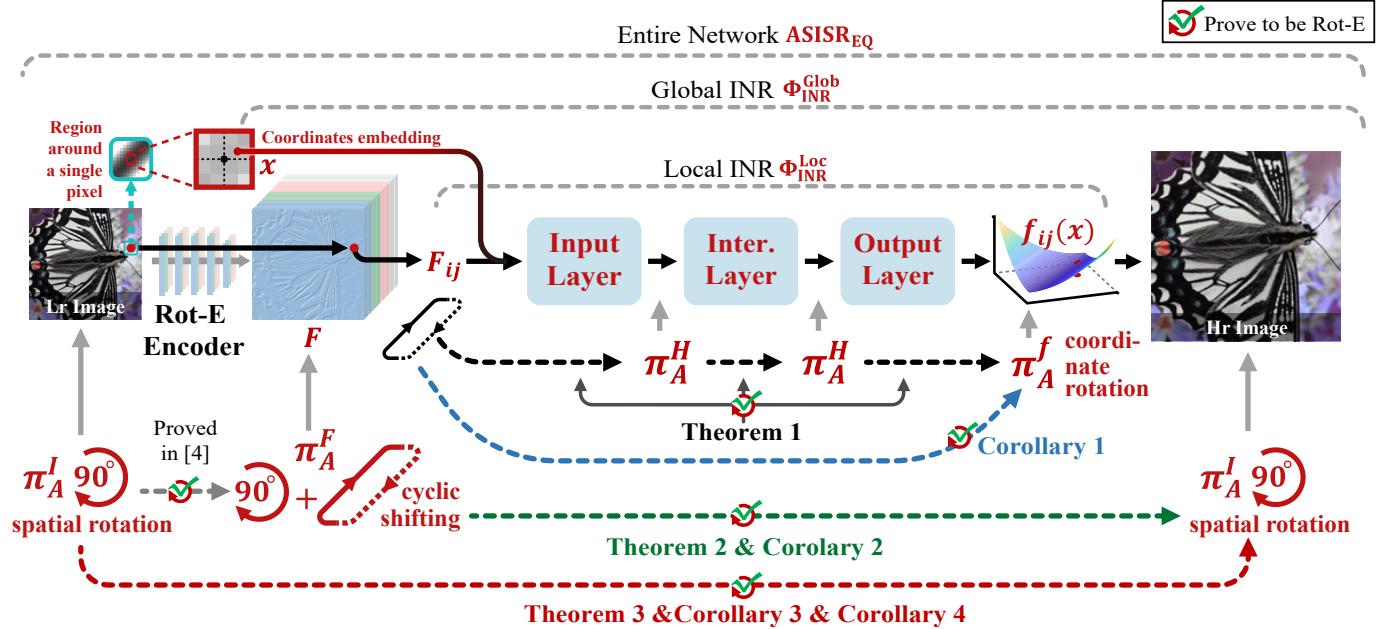


Fig. 1. Illustration of the proposed theoretical conclusions, which shows the correspondence between each theoretical conclusion and its respective network module.

As shown in Fig. 1, Theorem 1 shows that each of the proposed modules is theoretically rotation equivariant with the rotation defined in (7) and (8). Corollary 1 further shows that any INR constructed with the proposed modules is Rot-E on local implicit image function.

Theorem 2 and Corollary 2 provide the rotation equivariance analysis of the proposed INR on the entire image. The error bounds in the general case and the conclusions in simple special cases are presented. Particularly, Theorem 2 shows that the rotation equivariance error of the proposed INR mainly depends on the mesh size and the partial derivative of INR. When the mesh size of the LR image approaches zero, the equivariant error will also approach zero.

Finally, Theorem 3, Corollary 3 and Corollary 4 provide the rotation equivariance analysis of the proposed INR on the entire network that constructed by the commonly used Rot-E encoder and the proposed Rot-E INR.

**Examples for intuitively demonstrating the function of the theoretical results.** In practice, the most common choice is to set  $t = 4$ , i.e., to make the network  $\frac{\pi}{2}$  rotation-equivariant. In this case, when rotating the input image  $\frac{k\pi}{2}$  degree, where  $k$  is an integer, the feature map outputted by the Rot-E encoder,  $F$ , will be rotated for  $\frac{k\pi}{2}$  degree, and cyclic shifted for  $k$  times along its group dimension. For any selected pixel,  $F_{ij}$ , it will be cyclic shifted for  $k$  times along its group dimension (since the spatial rotation at a single pixel point would cause no effect). Then, by Theorem 1, we can find that all the feature function in INR ( $H_{ij}(x, A)$ ), will correspondingly undergo a  $\frac{k\pi}{2}$  rotation in coordinates be cyclic shifted in the group dimension for  $k$  times. Furthermore, by Corollary 1, we know that the local implicit image function constructed with the proposed INR framework can coordinately rotate for  $\frac{k\pi}{2}$  degree. By Corollary 2, we can then know that the output high-resolution-image will rotate for  $\frac{k\pi}{2}$  degree, according to the rotation and cyclic shifting on  $F$ . Finally, by Corollary 4, we know that when the input image is rotated for  $\frac{k\pi}{2}$  degree, the output image will also rotate accordingly.

Notably, the conclusions of the above examples apply not only to the entire image but also to the local structures of any image, because the ASISR network is translation equivariant.

#### 4 PROOFS OF THEOREM 1 AND COROLLARY 1

**Theorem 1.** For any pixel position  $(i, j)$ , let  $F_{ij}$  denotes the feature provided by a Rot-E encoder with its equivariant group denoted as  $S$ ; let  $W_{in}$ ,  $W$  and  $W_{out}$  be the parameters defined in (9), (10) and (11), respectively; and let  $H_{ij}$  denotes the feature function achieved by proposed input layer or intermediate layer, respectively, then the following results are satisfied for  $\forall \tilde{A} \in S$ :

$$\begin{aligned} W_{in} \hat{\diamond}_{\varphi} \left( \pi_{\tilde{A}}^F (F_{ij}) \right) &= \pi_{\tilde{A}}^H [W_{in} \hat{\diamond}_{\varphi} F_{ij}], \\ W \hat{\circ} \left( \pi_{\tilde{A}}^H [H_{ij}] \right) &= \pi_{\tilde{A}}^H [W \hat{\circ} H_{ij}], \\ W_{out} \check{\circ}_{\psi} \left( \pi_{\tilde{A}}^H [H_{ij}] \right) &= \pi_{\tilde{A}}^f [W_{out} \check{\circ}_{\psi} H_{ij}], \end{aligned} \quad (12)$$

where  $\pi_{\tilde{A}}^F$ ,  $\pi_{\tilde{A}}^H$  and  $\pi_{\tilde{A}}^f$  are defined in (7), (8).

*Proof.* 1) With the definition of the input layer for Rot-E INR, let  $C = \tilde{A}^{-1}A$ , we can deduce that for  $\forall x \in \mathbb{R}^2$  and  $B \in S$ :

$$\begin{aligned} \left[ W_{in} \hat{\diamond}_{\varphi} \left( \pi_{\tilde{A}}^F (F_{ij}) \right) \right] (x, B) &= \sum_{A \in S} \varphi \left( W_{in}^{B^{-1}A}, F_{ij}^{\tilde{A}^{-1}A}, A^{-1}x \right) \\ &= \sum_{C \in S} \varphi \left( W_{in}^{B^{-1}\tilde{A}C}, F_{ij}^C, (\tilde{A}C)^{-1}x \right) \\ &= \sum_{C \in S} \varphi \left( W_{in}^{(\tilde{A}^{-1}B)^{-1}C}, F_{ij}^C, C^{-1}(\tilde{A}^{-1}x) \right) \\ &= [W_{in} \hat{\diamond}_{\varphi} F_{ij}] (\tilde{A}^{-1}x, \tilde{A}^{-1}B) = \pi_{\tilde{A}}^H [W_{in} \hat{\diamond}_{\varphi} F_{ij}] (x, B). \end{aligned} \quad (13)$$

This proves the conclusion of the input layer.

2) With the definition of the intermediate layer for Rot-E INR, let  $C = \tilde{A}^{-1}B$ , we can deduce that for  $\forall x \in \mathbb{R}^2$  and  $A \in S$ :

$$\begin{aligned} \left[ W \hat{\circ} \left( \pi_{\tilde{A}}^H [H_{ij}] \right) \right] (x, A) &= \sum_{B \in S} W^{A^{-1}B} \cdot H_{ij} (\tilde{A}^{-1}x, \tilde{A}^{-1}B) \\ &= \sum_{C \in S} W^{A^{-1}\tilde{A}C} \cdot H_{ij} (\tilde{A}^{-1}x, C) \\ &= \sum_{C \in S} W^{(\tilde{A}^{-1}A)^{-1}C} \cdot H_{ij} (\tilde{A}^{-1}x, C) \\ &= [W \hat{\circ} H_{ij}] (\tilde{A}^{-1}x, \tilde{A}^{-1}A) = \pi_{\tilde{A}}^H [W \hat{\circ} H_{ij}] (x, A). \end{aligned} \quad (14)$$

3) With the definition of the output layer for Rot-E INR, let  $C = \tilde{A}^{-1}A$ , we can deduce that for  $\forall x \in \mathbb{R}^2$ :

$$\begin{aligned} \left[ W_{out} \check{\circ}_{\psi} \left( \pi_{\tilde{A}}^H [H_{ij}] \right) \right] (x) &= \psi_{W_{out}^2} \left( \sum_{A \in S} W_{out}^1 \cdot \hat{H}_{ij} (\tilde{A}^{-1}x, \tilde{A}^{-1}A) \right) \\ &= \psi_{W_{out}^2} \left( \sum_{C \in S} W_{out}^1 \cdot \hat{H}_{ij} (\tilde{A}^{-1}x, C) \right) \\ &= [W_{out} \check{\circ}_{\psi} H_{ij}] (\tilde{A}^{-1}x) = \pi_{\tilde{A}}^f [W_{out} \check{\circ}_{\psi} H_{ij}] (x). \end{aligned} \quad (15)$$

These deductions complete the proof.  $\square$

**Corollary 1.** For any INR constructed by the proposed modules, denote its local operation as  $\Phi_{\text{INR}}^{\text{Loc}}$ . Then, for arbitrary pixel position  $(i, j)$ , let  $S$  denotes the equivariant group, and the following result is satisfied for any  $\tilde{A} \in S$ :

$$\Phi_{\text{INR}}^{\text{Loc}}(\pi_{\tilde{A}}^F(F_{ij})) = \pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Loc}}(F_{ij})]. \quad (16)$$

*Proof.* With the conclusions in Theorem 1, it's easy to deduce that

$$\begin{aligned} \Phi_{\text{INR}}^{\text{Loc}}(\pi_{\tilde{A}}^F(F_{ij})) &= W_{\text{out}} \circ \psi W_1 \circ \dots \circ W_L \circ W_{\text{in}} \circ \hat{\phi}_{\varphi}(\pi_{\tilde{A}}^F(F_{ij})) \\ &= W_{\text{out}} \circ W_1 \circ \dots \circ W_L \circ (\pi_{\tilde{A}}^H[W_{\text{in}} \circ \hat{\phi}_{\varphi}(F_{ij})]) \\ &= W_{\text{out}} \circ W_1 \circ \dots \circ (\pi_{\tilde{A}}^H[W_L \circ W_{\text{in}} \circ \hat{\phi}_{\varphi}(F_{ij})]) \\ &= W_{\text{out}} \circ (\pi_{\tilde{A}}^H[W_1 \circ \dots \circ W_L \circ W_{\text{in}} \circ \hat{\phi}_{\varphi}(F_{ij})]) \\ &= \pi_{\tilde{A}}^f[W_{\text{out}} \circ W_1 \circ \dots \circ W_L \circ W_{\text{in}} \circ \hat{\phi}_{\varphi}(F_{ij})] \\ &= \pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Loc}}(F_{ij})], \end{aligned} \quad (17)$$

which complete the proof.  $\square$

## 5 PROOFS OF THEOREM 2 AND COROLLARY 2

**Theorem 2.** Assume the output feature map of a Rot-E encoder, denoted as  $F$ , is discredited from smooth function  $e : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$  whose channel number is  $n$ , and the channel number of the SR image is  $n_0$ . Denote the equivariant group as  $S$  with  $|S| = t$ , and denote the global operator corresponding to the INR constructed with the proposed modules as  $\Phi_{\text{INR}}^{\text{Glob}}$ . If the following conditions are satisfied for  $\forall x \in \mathbb{R}^2$ ,  $A \in S$ ,  $c = 1, 2, \dots, \frac{n}{t}$  and  $c_0 = 1, 2, \dots, n_0$ :

$$\|\nabla_x e_c(x, A)\| \leq G_e, \quad \left\| \nabla_x (\Phi_{\text{INR}}^{\text{Glob}}(F)(x))_{c_0} \right\| \leq G_{\text{INR}}^x, \quad \sup_{i,j} \left\{ \left\| \nabla_{F_{ij}} (\Phi_{\text{INR}}^{\text{Glob}}(F)(x))_{c_0} \right\| \right\} \leq G_{\text{INR}}^F, \quad (18)$$

then, for  $\forall \tilde{A} \in S$  and  $x \in \mathbb{R}^2$ , the following results is satisfied<sup>1</sup>:

$$\left| \Phi_{\text{INR}}^{\text{Glob}}(\pi_{\tilde{A}}^F(F))(x) - \pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Glob}}(F)](x) \right| \leq C\delta, \quad (19)$$

where  $C = \sqrt{2n}G_{\text{INR}}^F G_e + \sqrt{2}G_{\text{INR}}^x$  and  $\delta$  is the mesh size of the LR image.

*Proof.* Since  $\Phi_{\text{INR}}^{\text{Glob}}(F)(x) = \Phi_{\text{INR}}^{\text{Loc}}(F_{\hat{i}\hat{j}})(x - x_{\hat{i}\hat{j}})$ , where  $(\hat{i}, \hat{j})$  is the position of the pixel nearest to  $x$ , then we can deduce that for  $\forall \tilde{A} \in S$ ,

$$\Phi_{\text{INR}}^{\text{Glob}}(\pi_{\tilde{A}}^F(F))(x) = \Phi_{\text{INR}}^{\text{Loc}}\left(\left(\pi_{\tilde{A}}^F(F)\right)_{\hat{i}\hat{j}}\right)(x - x_{\hat{i}\hat{j}}) = \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(\tilde{A}^{-1}x_{\hat{i}\hat{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - x_{\hat{i}\hat{j}}). \quad (20)$$

Beside, we have

$$\pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Glob}}(F)](x) = \Phi_{\text{INR}}^{\text{Glob}}(F)(\tilde{A}^{-1}x) = \Phi_{\text{INR}}^{\text{Loc}}(F_{\tilde{i}\tilde{j}})(\tilde{A}^{-1}x - x_{\tilde{i}\tilde{j}}), \quad (21)$$

where  $(\tilde{i}, \tilde{j})$  is the position of the pixel nearest to  $\tilde{A}^{-1}x$ . Besides, with Corollary 1, we can deduce that for  $\forall \hat{x} \in \mathbb{R}^2$ ,

$$\begin{aligned} \Phi_{\text{INR}}^{\text{Loc}}(F_{\tilde{i}\tilde{j}})(\hat{x}) &= \pi_{\tilde{A}^{-1}}^f[\pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Loc}}(F_{\tilde{i}\tilde{j}})]](\hat{x}) \\ &= \pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Loc}}(F_{\tilde{i}\tilde{j}})](\tilde{A}\hat{x}) \\ &= \Phi_{\text{INR}}^{\text{Loc}}(\pi_{\tilde{A}}^F(F_{\tilde{i}\tilde{j}}))(\tilde{A}\hat{x}) \\ &= \Phi_{\text{INR}}^{\text{Loc}}\left(\left[F_{\tilde{i}\tilde{j}}^{\tilde{A}^{-1}A_k}\right]_{k=1}^t\right)(\tilde{A}\hat{x}) \\ &= \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(x_{\tilde{i}\tilde{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(\tilde{A}\hat{x}), \end{aligned} \quad (22)$$

where we have apply Corollary 1 in the third line. Then, let  $\hat{x} = \tilde{A}^{-1}x - x_{\tilde{i}\tilde{j}}$ , we can deduce

$$\pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Glob}}(F)](x) = \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(x_{\tilde{i}\tilde{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - \tilde{A}x_{\tilde{i}\tilde{j}}). \quad (23)$$

<sup>1</sup>for a tensor  $X$ ,  $|X| \leq \epsilon$  means all elements in  $X$  is not greater than  $\epsilon$ .

Since  $\delta$  is the mesh size of the LR image, then it is also easy to find that  $\|x_{\tilde{i}\tilde{j}} - x\| \leq \frac{\sqrt{2}}{2}\delta$  and  $\|x_{\tilde{i}\tilde{j}} - \tilde{A}^{-1}x\| \leq \frac{\sqrt{2}}{2}\delta$ . Then we can deduce that

$$\|\tilde{A}^{-1}x_{\tilde{i}\tilde{j}} - x_{\tilde{i}\tilde{j}}\| = \|x_{\tilde{i}\tilde{j}} - \tilde{A}x_{\tilde{i}\tilde{j}}\| \leq \|x_{\tilde{i}\tilde{j}} - x\| + \|\tilde{A}x_{\tilde{i}\tilde{j}} - x\| = \|x_{\tilde{i}\tilde{j}} - x\| + \|x_{\tilde{i}\tilde{j}} - \tilde{A}^{-1}x\| \leq \frac{\sqrt{2}}{2}\delta + \frac{\sqrt{2}}{2}\delta = \sqrt{2}\delta. \quad (24)$$

Then, combining (20) and (23), we have

$$\begin{aligned} & \left| \Phi_{\text{INR}}^{\text{Glob}}(\pi_{\tilde{A}}^F(F))(x) - \pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Glob}}(F)](x) \right| \\ &= \left| \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(\tilde{A}^{-1}x_{\tilde{i}\tilde{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - x_{\tilde{i}\tilde{j}}) - \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(x_{\tilde{i}\tilde{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - \tilde{A}x_{\tilde{i}\tilde{j}}) \right| \\ &\leq \left| \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(\tilde{A}^{-1}x_{\tilde{i}\tilde{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - x_{\tilde{i}\tilde{j}}) - \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(x_{\tilde{i}\tilde{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - x_{\tilde{i}\tilde{j}}) \right| + \\ &\quad \left| \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(x_{\tilde{i}\tilde{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - x_{\tilde{i}\tilde{j}}) - \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(x_{\tilde{i}\tilde{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - \tilde{A}x_{\tilde{i}\tilde{j}}) \right|, \\ &\leq \sup_{A,x,c_0} \left\{ \left\| \nabla_{\hat{x}} \left( \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e(\hat{x}, A_k)\right]_{k=1}^t\right)(x) \right)_{c_0} \right\| \right\} \|\tilde{A}^{-1}x_{\tilde{i}\tilde{j}} - x_{\tilde{i}\tilde{j}}\| + \sup_{F,c_0,i,j} \left\{ \left\| \nabla_x \left( \Phi_{\text{INR}}^{\text{Loc}}(F_{ij})(x) \right)_{c_0} \right\| \right\} \|x_{\tilde{i}\tilde{j}} - \tilde{A}x_{\tilde{i}\tilde{j}}\| \\ &\leq \sup_{A,x,c_0} \left\{ \left\| \nabla_{\hat{x}} \left( \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e(\hat{x}, A_k)\right]_{k=1}^t\right)(x) \right)_{c_0} \right\| \right\} \sqrt{2}\delta + \sup_{F,c_0} \left\{ \left\| \nabla_x \left( \Phi_{\text{INR}}^{\text{Glob}}(F)(x) \right)_{c_0} \right\| \right\} \sqrt{2}\delta \\ &\leq \sup_{A,x,c_0,i,j} \left\{ \left\| \nabla_{F_{ij}} \left( \Phi_{\text{INR}}^{\text{Glob}}(F)(x) \right)_{c_0} \right\| \cdot \left\| \nabla_{\hat{x}} \left[ [e_c(\hat{x}, A_k)]_{k=1}^t \right]_{c=1}^{\frac{n}{t}} \right\| \right\} \sqrt{2}\delta + \sqrt{2}G_{\text{INR}}^x\delta \\ &\leq \sup_{x,c_0,i,j} \left\{ \left\| \nabla_{F_{ij}} \left( \Phi_{\text{INR}}^{\text{Glob}}(F)(x) \right)_{c_0} \right\| \cdot \sup_A \left\{ \left\| [\nabla_{\hat{x}} e_c(\hat{x}, A_k)]_{k=1}^t \right\|_{c=1}^{\frac{n}{t}} \right\} \right\} \sqrt{2}\delta + \sqrt{2}G_{\text{INR}}^x\delta \\ &\leq \sup_{x,c_0,i,j} \left\{ \left\| \nabla_{F_{ij}} \left( \Phi_{\text{INR}}^{\text{Glob}}(F)(x) \right)_{c_0} \right\| \cdot \sqrt{n} \sup_A \{ \|\nabla_{\hat{x}} e_c(\hat{x}, A_k)\| \} \right\} \sqrt{2}\delta + \sqrt{2}G_{\text{INR}}^x\delta \\ &\leq \sqrt{2n}G_{\text{INR}}^F G_e \delta + \sqrt{2}G_{\text{INR}}^x\delta = \left( \sqrt{2nt}G_{\text{INR}}^F G_e + \sqrt{2}G_{\text{INR}}^x \right) \delta. \end{aligned} \quad (25)$$

This prove the conclusion.  $\square$

**Corollary 2.** Under the same condition of Theorem 2, when  $t = 2$  or  $t = 4$ , the following result hold: for any  $\tilde{A} \in S$ :

$$\Phi_{\text{INR}}^{\text{Glob}}(\pi_{\tilde{A}}^F(F)) = \pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Glob}}(F)]. \quad (26)$$

*Proof.* When  $t = 2$  or  $t = 4$ ,  $\forall \tilde{A} \in S$ , its corresponding rotation degree would be  $\frac{k\pi}{2}$ , where  $k$  is a integer. Then, for  $\forall x$ , let  $(\hat{i}, \hat{j})$  and  $(\tilde{i}, \tilde{j})$  be the position of the pixels nearest to  $x$  and  $\tilde{A}^{-1}x$  respectively, it's easy to find that

$$x_{\hat{i}\hat{j}} = \tilde{A}x_{\tilde{i}\tilde{j}}. \quad (27)$$

By Eq. (20) and (23), we can find that

$$\begin{aligned} & \left| \Phi_{\text{INR}}^{\text{Glob}}(\pi_{\tilde{A}}^F(F))(x) - \pi_{\tilde{A}}^f[\Phi_{\text{INR}}^{\text{Glob}}(F)](x) \right| \\ &= \left| \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(\tilde{A}^{-1}x_{\hat{i}\hat{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - x_{\hat{i}\hat{j}}) - \Phi_{\text{INR}}^{\text{Loc}}\left(\left[e\left(x_{\tilde{i}\tilde{j}}, \tilde{A}^{-1}A_k\right)\right]_{k=1}^t\right)(x - \tilde{A}x_{\tilde{i}\tilde{j}}) \right| = 0. \end{aligned} \quad (28)$$

This prove the conclusion.  $\square$

## 6 PROOFS OF THEOREM 3 AND COROLLARY 3

Since the Rot-E encoder is based on the Rot-E convolutions proposed in [3], we need to first introduce the definition of rotation equivariant convolutions before prove the theorem 3.

**Input layer.** Suppose the filter is of size  $p \times p$ , then, the mesh grids for filter and image can be respectively represented as:

$$y_{ij} = \left( \left( i - \frac{p+1}{2} \right) \delta, \left( j - \frac{p+1}{2} \right) \delta \right)^T, i, j = 1, 2, \dots, p. \quad (29)$$

The filter of the input multi-channel convolution layer can be represented as

$$\Psi_{ij}^{c,d,A} = \phi_{cd}(A^{-1}y_{ij}), \quad (30)$$

where  $\phi_{cd}$  is the parameterized filter,  $A \in S$ ,  $c = 1, 2, \dots, n_c$ ,  $d = 1, 2, \dots, n_d$ ,  $n_c$  and  $n_d$  are the input and output channel numbers, respectively. Denoting multi-channel convolution of  $\Psi$  and  $I$  in the input layer as  $F = \Psi \star I$ , then it can be calculated by

$$(\Psi \star I)^{d,A} = \sum_c \Psi^{c,d,A} * I^c, \quad (31)$$

where  $*$  denotes the 2-D convolution operation. It can be also rewritten in the following more detailed formulation:

$$\hat{e}_d(x_{ij}, A) = \sum_{c,y \in \Lambda} \phi_{cd}(A^{-1}y) r_c(x_{ij} - y), \quad (32)$$

where  $\Lambda$  is a set of indexes, denoted as  $\Lambda = \{y_{\hat{j}} | \hat{i}, \hat{j} = 1, 2, \dots, p\}$ ,  $A \in S$ ,  $i = 1, 2, \dots, h$  and  $j = 1, 2, \dots, w$ .

**Intermediate layer.** The filter of the intermediate multi-channel convolution layer can be represented as

$$\Phi_{ij}^{c,d,A,B} = \phi_{Ac}(B^{-1}y_{ij}), \quad (33)$$

where  $\phi_{Ac}$  is the parameterized filter,  $A, B \in S$ ,  $c = 1, 2, \dots, n_c$ ,  $d = 1, 2, \dots, n_d$ ,  $n_c$  and  $n_d$  are the input and output channel numbers, respectively. Denoting the multi-channel convolution of  $\Phi$  and  $F$  in the intermediate layer as  $\hat{F} = \hat{\Phi} \star F$ , then it can be calculated by

$$(\hat{\Phi} \star F)^{d,B} = \sum_{c,A} \Phi_{ij}^{c,d,A,B} * F^{c,A}. \quad (34)$$

It can also be rewritten in the following more detailed formulation:

$$\hat{e}_d(x_{ij}, B) = \sum_{c,A,y \in \Lambda} \phi_{Ac}(B^{-1}y) e_c(x_{ij} - y, BA). \quad (35)$$

We also need to introduce the following lemmas.

**Lemma 1** (Xie. 2022 [3]). Assume that an image  $I$  is discretized from the smooth function  $r : \mathbb{R}^2 \rightarrow \mathbb{R}$  by (1), a feature map  $F$  is discretized from the smooth function  $e : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$  by (2),  $|S| = t$ , and filters  $\Psi, \Phi$  and  $\tilde{\Delta}$  are generated from  $\phi_{in}$ ,  $\phi_{out}$  and  $\phi_A$ ,  $\forall A \in S$ , respectively. For any  $A \in S$ ,  $x \in \mathbb{R}^2$ , the following conditions are satisfied:

$$\begin{aligned} & |r(x)|, |e(x, A)| \leq F_1, \|\nabla_x r(x)\|, \|\nabla_x e(x, A)\| \leq G_1, \|\nabla_x^2 r(x)\|, \|\nabla_x^2 e(x, A)\| \leq H_1, \\ & |\phi_{in}(x)|, |\phi_A(x)| \leq F_2, \|\nabla_x \phi_{in}(x)\|, \|\nabla_x \phi_A(x)\|, \|\nabla_x^2 \phi_{in}(x)\|, \|\nabla_x^2 \phi_A(x)\| \leq H_2, \\ & \forall \|x\| \geq (p+1)\delta/2, \phi_{in}(x), \phi_A(x) = 0, \end{aligned} \quad (36)$$

where  $p$  is the filter size,  $\delta$  is the mesh size, and  $\nabla_x$  and  $\nabla_x^2$  denote the operators of gradient and Hessian matrix with respect to  $x$ , respectively, then for any  $\tilde{A} \in S$ , the following results are satisfied:

$$|\Psi \star \pi_{\tilde{A}}^I(I) - \pi_{\tilde{A}}^F(\Psi \star I)| \leq \frac{C}{2}(p+1)^2\delta^2, \quad \left| \Phi \star \pi_{\tilde{A}}^F(F) - \pi_{\tilde{A}}^F(\Phi \star F) \right| \leq \frac{C}{2}(p+1)^2\delta^2t, \quad (37)$$

where  $C = F_1H_2 + F_2H_1 + 2G_1G_2$ ;  $\Psi$  and  $\Phi$  are defined by (31) and (34), respectively;  $\pi_{\tilde{A}}^I$  and  $\pi_{\tilde{A}}^F$  are defined in (5); and  $\|\cdot\|_\infty$  represents the infinity norm.

**Lemma 2** (Fu. 2023 [4]). For an image  $I$  with size  $h \times w \times c$ , and a  $L$ -layer rotation equivariant CNN network  $\text{CNN}_{eq}(\cdot)$ , whose channel number of the  $i^{th}$  layer is  $L_i$ , rotation equivariant subgroup is  $S \leqslant O(2)$ ,  $|S| = t$ , and activation function is set as ReLU. If the latent continuous function of the  $c^{th}$  channel of  $I$  denoted as  $r_c : \mathbb{R}^2 \rightarrow \mathbb{R}$ , and the latent continuous function of any convolution filters in the  $i^{th}$  layer denoted as  $\phi^l : \mathbb{R}^2 \rightarrow \mathbb{R}$ , where  $l \in \{1, \dots, L\}$ ,  $c \in \{1, \dots, C\}$ , for any  $x \in \mathbb{R}^2$ , the following conditions are satisfied:

$$\begin{aligned} & |r_c(y)| \leq F_0, \|\nabla_x r_c(y)\| \leq G_0, \|\nabla_x^2 r_c(y)\| \leq H_0, \\ & |\phi^l(y)| \leq F_l, \|\nabla_x \phi^l(y)\| \leq G_l, \|\nabla_x^2 \phi^l(y)\| \leq H_l, \\ & \forall \|x\| \geq (p+1)\delta/2, \phi^l(y) = 0, \end{aligned} \quad (38)$$

where  $p$  is the filter size,  $\delta$  is the mesh size, and  $\nabla_x$  and  $\nabla_x^2$  denote the operators of gradient and Hessian matrix, respectively. Denote

$$e_d^l(y, B) = \begin{cases} \sum_{c,y \in \Lambda} \phi_{cd}^1(B^{-1}y) r_c(y - y) & \text{if } l = 1, \\ \sum_{c,A,y \in \Lambda} \phi_{Ac}^l(B^{-1}y) e_c^{l-1}(y - y, BA) & \text{if } l \neq 1, L \end{cases} \quad (39)$$

where  $\Lambda = \left\{ \left( \left( a - \frac{p+1}{2} \right) \delta, \left( b - \frac{p+1}{2} \right) \delta \right)^T | a, b = 1, 2, \dots, p \right\}$ ,  $\phi_{cd}^1$  and  $\phi_{Ac}^l$  are filters in the first layer and other layers respectively. Then, for  $\forall B \in S$  the following results are satisfied:

$$|e_d^l(y, B)| \leq F_0 \mathcal{F}_l, \quad (40)$$

$$\left| \nabla_x e_d^l(y, B) \right| \leq \left( \sum_{m=1}^l \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_l, \quad (41)$$

$$\left| \nabla_x^2 e_d^l(y, B) \right| \leq \sum_{m=1}^l \left( \frac{H_m F_0}{F_m} + 2 \sum_{n=1}^{m-1} \frac{G_m G_n F_0}{F_m F_n} + 2 \frac{G_m G_0}{F_m} + \frac{H_0}{l} \right) \mathcal{F}_l, \quad (42)$$

where  $\mathcal{F}_l = \prod_{k=1}^l n_{k-1} p^2 F_k$ ,  $\forall l = 1, 2, \dots, L-1$ .

Moreover it is also necessary to prove the following lemma.

**Lemma 3.** For an image  $I$  with size  $h \times w \times n_0$ , and a Rot-E encoder constructed by  $L$ -layer rotation equivariant CNN network,  $\Phi_{\text{CNN}}^{\text{EQ}}(\cdot)$ , whose channel number of the  $l^{\text{th}}$  layer is  $n_l$ , rotation equivariant subgroup is  $S \leq O(2)$ ,  $|S| = t$ , and activation function is set as ReLU. If the latent continuous function of the  $c^{\text{th}}$  channel of  $I$  denoted as  $r_c : \mathbb{R}^2 \rightarrow \mathbb{R}$ , and the latent continuous function of any convolution filters in the  $l^{\text{th}}$  layer denoted as  $\phi^l : \mathbb{R}^2 \rightarrow \mathbb{R}$ , for any  $x \in \mathbb{R}^2$ ,  $l \in \{1, \dots, L\}$ ,  $c \in \{1, \dots, n_0\}$ , the following conditions are satisfied:

$$\begin{aligned} |r_c(x)| &\leq F_0, \|\nabla_x r_c(x)\| \leq G_0, \|\nabla_x^2 r_c(x)\| \leq H_0, \\ |\phi^l(x)| &\leq F_l, \|\nabla_x \phi^l(x)\| \leq G_l, \|\nabla_x^2 \phi^l(x)\| \leq H_l, \\ \forall \|x\| &\geq (p+1)\delta/2, \phi_l(x) = 0, \end{aligned} \quad (43)$$

where  $p$  is the filter size,  $\delta$  is the mesh size, and  $\nabla_x$  and  $\nabla_x^2$  denote the operators of gradient and Hessian matrix, respectively. For  $\forall \tilde{A} \in S$ :

$$\left| \Phi_{\text{CNN}}^{\text{EQ}} \left[ \pi_{\tilde{A}}^I \right] (I) - \pi_{\tilde{A}}^F \left[ \Phi_{\text{CNN}}^{\text{EQ}} \right] (I) \right| \leq C \delta^2, \quad (44)$$

where

$$C = 2\mathcal{F} \sum_{l=1}^L \sum_{m=1}^l \left( \frac{H_m F_0}{F_m} + 2 \sum_{n=1}^{m-1} \frac{G_m G_n F_0}{F_m F_n} + 2 \frac{G_m G_0}{F_m} + \frac{H_0}{l} \right), \quad (45)$$

and  $\mathcal{F} = \prod_{l=1}^L n_{l-1} p^2 F_l$ .

*Proof.* A  $L$ -layer rotation equivariant CNN based encoder includes 1 input layer and  $L-1$  intermediate layers, which can be formally defined as

$$\Phi_{\text{CNN}}^{\text{EQ}}(I) = \Phi_L \star \dots \star \Phi_l \star \dots \star \Phi_2 \star \Psi \star (I), \quad (46)$$

where  $I$  is a input image. It should be noted that the ReLU activation does not amplify equivariant errors, thus for the sake of simplicity, we did not include ReLU in the proof, but it is easy to see that the proof carries over to the case with ReLU as well. Then,

$$\begin{aligned} &\left| \Phi_{\text{CNN}}^{\text{EQ}} \left[ \pi_{\tilde{A}}^I \right] (I) - \pi_{\tilde{A}}^F \left[ \Phi_{\text{CNN}}^{\text{EQ}} \right] (I) \right| \\ &= \left| \Phi_L \star \dots \star \Phi_l \star \dots \star \Phi_2 \star \Psi \star \pi_{\tilde{A}}^I (I) - \pi_{\tilde{A}}^F (\Phi_L \star \dots \star \Phi_l \star \dots \star \Phi_2 \star \Psi \star I) \right| \\ &\leq \left| \Phi_L \star \dots \star \Phi_l \star \dots \star \Phi_2 \star \Psi \star \pi_{\tilde{A}}^I (I) - \Phi_L \star \dots \star \Phi_l \star \dots \star \Phi_2 \star \pi_{\tilde{A}}^F (\Psi \star I) \right| + \\ &\quad \sum_{l=2}^L \left| \Phi_L \star \dots \star \Phi_l \star \pi_{\tilde{A}}^F (\dots \star \Phi_2 \star \Psi \star I) - \Phi_L \star \dots \star \pi_{\tilde{A}}^F (\Phi_l \star \dots \star \Phi_2 \star \Psi \star I) \right|. \end{aligned} \quad (47)$$

We denote  $y_1, y_l$  ( $l = 2, 3, \dots, L-1$ ), and  $y_L$  as the filter indexes of the input layer,  $l^{\text{th}}$  intermediate layer and output layer, respectively, let  $x$  define the coordinate of the  $(i, j)$  position.

1) For the input layer, with (32) and (35), let  $x$  denote the coordinate of position  $(i, j)$ ,  $\Lambda$  defined as (32), then we have:

$$\begin{aligned} &\left| \left( \Phi_L \star \dots \star \Phi_l \star \dots \star \Phi_2 \star \Psi \star \pi_{\tilde{A}}^I (I) - \Phi_L \star \dots \star \Phi_l \star \dots \star \Phi_2 \star \pi_{\tilde{A}}^F (\Psi \star I) \right)_{ij}^{c_L} \right| \\ &= \left| \sum_{\substack{c_{L-1} \in S \\ A_{L-1} \in S \\ y_L \in \Lambda}} \dots \sum_{\substack{c_1 \\ A_1 \in S}} \sum_{\substack{c_0 \\ y_1 \in \Lambda}} \phi_{A_{L-1} c_{L-1} c_L}^L (A_L^{-1} y_L) \dots \phi_{A_1 c_1 c_2}^2 (A_2^{-1} y_2) \phi_{c_0 c_1}^1 (A_1^{-1} y_1) r_{c_0} (\tilde{A}^{-1} (x - y_L - \dots - y_2 - y_1)) \right. \\ &\quad \left. - \sum_{\substack{c_{L-1} \in S \\ A_{L-1} \in S \\ y_L \in \Lambda}} \dots \sum_{\substack{c_1 \\ A_1 \in S}} \sum_{\substack{c_0 \\ y_1 \in \Lambda}} \phi_{A_{L-1} c_{L-1} c_L}^L (A_L^{-1} y_L) \dots \phi_{A_1 c_1 c_2}^2 (A_2^{-1} y_2) \phi_{c_0 c_1}^1 (A_1^{-1} \tilde{A} y_1) r_{c_0} (\tilde{A}^{-1} (x - y_L - \dots - y_2) - y_1) \right| \end{aligned} \quad (48)$$

$$\begin{aligned}
&\leq \left| \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_1 \\ A_1 \in S \\ y_2 \in \Lambda}} \sum_{c_0} \phi_{A_{L-1} c_{L-1} c_L}^L(A_L^{-1} y_L) \cdots \phi_{A_1 c_1 c_2}^2(A_2^{-1} y_2) \right. \\
&\quad \left. \left( \sum_{y_1 \in \Lambda} \phi_{c_0 c_1}^1(A_1^{-1} y_1) r(\tilde{A}^{-1}(x - y_L - \cdots - y_2 - y_1)) - \sum_{y_1 \in \Lambda} \phi_{c_0 c_1}^1(A_1^{-1} \tilde{A} y_1) r_{c_0}(\tilde{A}^{-1}(x - y_L - \cdots - y_2) - y_1) \right) \right| \\
&\leq \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_1 \\ A_1 \in S \\ y_2 \in \Lambda}} \left| \sum_{c_0} \phi_{A_{L-1} c_{L-1} c_L}^L(A_L^{-1} y_L) \right| \cdots \left| \phi_{A_1 c_1 c_2}^2(A_2^{-1} y_2) \right| \\
&\quad \left| \sum_{y_1 \in \Lambda} \phi_{c_0 c_1}^1(A_1^{-1} y_1) r(\tilde{A}^{-1}(x - y_L - \cdots - y_2 - y_1)) - \sum_{y_1 \in \Lambda} \phi_{c_0 c_1}^1(A_1^{-1} \tilde{A} y_1) r_{c_0}(\tilde{A}^{-1}(x - y_L - \cdots - y_2) - y_1) \right| \\
&\leq \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_1 \\ A_1 \in S \\ y_2 \in \Lambda}} \sum_{c_0} F_L \cdots F_2 \\
&\quad \left| \sum_{y_1 \in \Lambda} \phi_{c_0 c_1}^1(A_1^{-1} y_1) r(\tilde{A}^{-1}(x - y_L - \cdots - y_2 - y_1)) - \sum_{y_1 \in \Lambda} \phi_{c_0 c_1}^1(A_1^{-1} \tilde{A} y_1) r_{c_0}(\tilde{A}^{-1}(x - y_L - \cdots - y_2) - y_1) \right|.
\end{aligned}$$

Let  $(\hat{i}, \hat{j})$  be the position in the image that corresponded to  $\hat{x} = x - y_L - y_{L-1} - \cdots - y_2$ , with Eq. (37) from Lemma 1 for the input layer, we can deduce the following result:

$$\begin{aligned}
&\left| \sum_{y_1 \in \Lambda} \phi_{c_0 c_1}^1(A_1^{-1} y_1) r(\tilde{A}^{-1}(x - y_L - \cdots - y_2 - y_1)) - \sum_{\substack{c_0 \\ y_1 \in \Lambda}} \phi_{c_0 c_1}^1(A_1^{-1} \tilde{A} y_1) r_{c_0}(\tilde{A}^{-1}(x - y_L - \cdots - y_2) - y_1) \right| \\
&= \left| \sum_{y_1 \in \Lambda} \phi_{c_0 c_1}^1(A_1^{-1} y_1) r(\tilde{A}^{-1}(\hat{x} - y_1)) - \sum_{\substack{c_0 \\ y_1 \in \Lambda}} \phi_{c_0 c_1}^1(A_1^{-1} \tilde{A} y_1) r_{c_0}(\tilde{A}^{-1}(\hat{x} - y_1)) \right| \\
&= \left| (\Psi \star \pi_{\tilde{A}}^I(I^c) - \pi_{\tilde{A}}^F(\Psi \star I^c))_{\hat{i}\hat{j}}^{A_1} \right| \leq \frac{C_1}{2}(p+1)^2 \delta^2,
\end{aligned} \tag{49}$$

and we have

$C_1 = \sup(\|\nabla_x^2 \phi^1(x)\| |r(x)| + |\phi^1(x)| \|\nabla_x^2 r(x)\| + 2 \|\nabla_x \phi^1(x)\| \|\nabla_x r(x)\|) \leq H_1 F_0 + F_1 H_0 + 2G_1 G_0$ . Therefore, combining (48) and (49), we have:

$$\begin{aligned}
&\left| \Phi_L \star \cdots \Phi_{l+1} \star \Phi_l \cdots \Phi_2 \star \Psi \star \pi_{\tilde{A}}^I(I) - \Phi_L \star \cdots \Phi_{l+1} \star \Phi_l \cdots \Phi_2 \star \pi_{\tilde{A}}^F(\Psi \star I) \right| \\
&\leq n_{L-1} p^2 F_L n_{L-2} p^2 F_{L-1} \cdots n_1 p^2 F_2 n_0 \frac{C_1}{2}(p+1)^2 \delta^2 \\
&\leq \left( \prod_{k=2}^L (n_{k-1} p^2 F_k) \right) n_0 \frac{C_1}{2}(p+1)^2 \delta^2 = 2\mathcal{F} \left( \frac{H_1}{F_1} F_0 + H_0 + 2 \frac{G_1}{F_1} G_0 \right) \delta^2.
\end{aligned} \tag{50}$$

2) For the other layers, i.e.,  $1 < l \leq L$ . Denoting the latent function of  $l$ -th layer feature as  $e^l(x, B)$  ( $B \in S$ ), and the feature map of  $l$ -th layer feature as  $F^l$ , then we have:

$$\begin{aligned}
&\left| \left( \Phi_L \star \cdots \Phi_l \star \pi_{\tilde{A}}^F(\cdots \Phi_2 \star \Psi \star I) - \Phi_L \star \cdots \pi_{\tilde{A}}^F(\Phi_l \star \cdots \Phi_2 \star \Psi \star I) \right)_{ij}^{c_L} \right| \\
&= \left| \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_{l-1} \\ A_{l-1} \in S \\ y_l \in \Lambda}} \phi_{A_{L-1} c_{L-1} c_L}^L(A_L^{-1} y_L) \cdots \phi_{A_{l-1} c_{l-1} c_l}^l(A_l^{-1} y_l) e_{c_l-1}^{l-1}(\tilde{A}^{-1}(x - y_L - \cdots - y_{l+1} - y_l), \tilde{A}^{-1} A_{l-1}) \right. \\
&\quad \left. - \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_{l-1} \\ A_{l-1} \in S \\ y_l \in \Lambda}} \phi_{A_{L-1} c_{L-1} c_L}^L(A_L^{-1} y_L) \cdots \phi_{A_{l-1} c_{l-1} c_l}^l(A_l^{-1} \tilde{A} y_l) e_{c_l-1}^{l-1}(\tilde{A}^{-1}(x - y_L - \cdots - y_{l+1}) - y_l, A_{l-1}) \right|
\end{aligned} \tag{51}$$

$$\begin{aligned}
&= \left| \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_l \\ A_l \in S \\ y_{l+1} \in \Lambda}} \sum_{c_{l-1}} \phi_{A_{L-1} c_{L-1} c_L}^L(A_L^{-1} y_L) \cdots \phi_{A_l c_l c_{l-1}}^{l+1}(A_{l+1}^{-1} y_{l+1}) \right. \\
&\quad \left( \sum_{\substack{A_{l-1} \in S \\ y_l \in \Lambda}} \phi_{A_{l-1} c_{l-1} c_l}^l(A_l^{-1} y_l) e_{c_l-1}^{l-1}(\tilde{A}^{-1}(x - y_L - \cdots - y_{l+1} - y_l), \tilde{A}^{-1} A_{l-1}) \right. \\
&\quad \left. - \sum_{\substack{A_{l-1} \in S \\ y_l \in \Lambda}} \phi_{A_{l-1} c_{l-1} c_l}^l(A_l^{-1} \tilde{A} y_l) e_{c_l-1}^{l-1}(\tilde{A}^{-1}(x - y_L - \cdots - y_{l+1}) - y_l, A_{l-1}) \right) \Bigg| \\
&\leq \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_l \\ A_l \in S \\ y_{l+1} \in \Lambda}} \sum_{c_{l-1}} \left| \phi_{A_{L-1} c_{L-1} c_L}^L(A_L^{-1} y_L) \right| \cdots \left| \phi_{A_l c_l c_{l-1}}^{l+1}(A_{l+1}^{-1} y_{l+1}) \right| \\
&\quad \left| \sum_{\substack{A_{l-1} \in S \\ y_l \in \Lambda}} \phi_{A_{l-1} c_{l-1} c_l}^l(A_l^{-1} y_l) e_{c_l-1}^{l-1}(\tilde{A}^{-1}(x - y_L - \cdots - y_{l+1} - y_l), \tilde{A}^{-1} A_{l-1}) \right. \\
&\quad \left. - \sum_{\substack{A_{l-1} \in S \\ y_l \in \Lambda}} \phi_{A_{l-1} c_{l-1} c_l}^l(A_l^{-1} \tilde{A} y_l) e_{c_l-1}^{l-1}(\tilde{A}^{-1}(x - y_L - \cdots - y_{l+1}) - y_l, A_{l-1}) \right| \\
&\leq \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_l \\ A_l \in S \\ y_{l+1} \in \Lambda}} \sum_{c_{l-1}} F_L \cdots F_{l+1} \left| \left( \pi_A^F(\Phi_l * F^l) - \Phi_l * \pi_A^F(F^l) \right)_{\hat{i}_l \hat{j}_l}^{A_l} \right| \\
&\leq \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_l \\ A_l \in S \\ y_{l+1} \in \Lambda}} \sum_{c_{l-1}} F_L \cdots F_{l+1} \frac{C_l}{2} (p+1)^2 \delta^2 = \left( \prod_{k=l+1}^L n_{k-1} p^2 F_k \right) n_{l-1} \frac{C_l}{2} (p+1)^2 \delta^2,
\end{aligned}$$

where  $(\hat{i}_l, \hat{j}_l)$  is the position that corresponded to  $\hat{x} = x - y_L - \cdots - y_{l+1}$ , and we adopt Eq. (37) from Lemma 1 in the last line, i.e.,

$$C_l = \sup \left( H_l \left| e_{c_l-1}^{l-1}(x, B) \right| + F_l \left\| \nabla_x^2 e_{c_l-1}^{l-1}(x, B) \right\| + 2G_l \left\| \nabla_x e_{c_l-1}^{l-1}(x, B) \right\| \right). \quad (52)$$

Substituting Eqs. (40), (41) and (42) into Eq. (52), then we have

$$\begin{aligned}
&\left| \Phi_L * \cdots * \Phi_l * \pi_A^F(\cdots * \Phi_2 * \Psi * I) - \Phi_L * \cdots * \pi_A^F(\Phi_l * \cdots * \Phi_2 * \Psi * I) \right| \\
&\leq \left( \prod_{k=l+1}^L n_{k-1} p^2 F_k \right) n_{l-1} \frac{(p+1)^2 \delta^2}{2} \left( H_l \left| e_{c_l-1}^{l-1}(x, B) \right| + F_l \left\| \nabla_x^2 e_{c_l-1}^{l-1}(x, B) \right\| + 2G_l \left\| \nabla_x e_{c_l-1}^{l-1}(x, B) \right\| \right) \\
&\leq 2 \left( \prod_{k=l}^L n_{k-1} p^2 F_k \right) \left( \frac{H_l}{F_l} \left| e_{c_l-1}^{l-1}(x, B) \right| + \left\| \nabla_x^2 e_{c_l-1}^{l-1}(x, B) \right\| + 2 \frac{G_l}{F_l} \left\| \nabla_x e_{c_l-1}^{l-1}(x, B) \right\| \right) \delta^2 \\
&\leq 2\mathcal{F} \left( \frac{H_l F_0}{F_l} + \sum_{m=1}^{l-1} \frac{H_m F_0}{F_m} + 2 \sum_{m=1}^{l-1} \sum_{n=1}^{m-1} \frac{G_m G_n F_0}{F_m F_n} + 2 \sum_{m=1}^{l-1} \frac{G_m G_0}{F_m} + H_0 + 2 \sum_{m=1}^{l-1} \frac{G_m G_l F_0}{F_m F_l} + 2 \frac{G_l G_0}{F_l} \right) \delta^2, \\
&\leq 2\mathcal{F} \left( \sum_{m=1}^l \frac{H_m F_0}{F_m} + 2 \sum_{m=1}^l \sum_{n=1}^{m-1} \frac{G_m G_n F_0}{F_m F_n} + 2 \sum_{m=1}^l \frac{G_m G_0}{F_m} + H_0 \right) \delta^2,
\end{aligned} \quad (53)$$

where we have used  $\frac{p+1}{p} \leq 2$  in the third line.

3) It is easy to find that (50) also satisfies the general term formula of (53). Substituting Eqs. (50) and (53) into Eq. (47), we can get:

$$\begin{aligned}
&\left| \Phi_{\text{CNN}}^{\text{EQ}} \left[ \pi_A^I \right] (I) - \pi_A^F \left[ \Phi_{\text{CNN}}^{\text{EQ}} \right] (I) \right| \leq 2\mathcal{F} \sum_{l=1}^L \left( \sum_{m=1}^l \frac{H_m F_0}{F_m} + 2 \sum_{m=1}^l \sum_{n=1}^{m-1} \frac{G_m G_n F_0}{F_m F_n} + 2 \sum_{m=1}^l \frac{G_m G_0}{F_m} + H_0 \right) \delta^2 \\
&= 2\mathcal{F} \sum_{l=1}^L \sum_{m=1}^l \left( \frac{H_m F_0}{F_m} + 2 \sum_{n=1}^{m-1} \frac{G_m G_n F_0}{F_m F_n} + 2 \frac{G_m G_0}{F_m} + \frac{H_0}{l} \right) \delta^2.
\end{aligned} \quad (54)$$

Then, we achieve Eq. (45).  $\square$

With above lemmas, we can now prove the Theorem 3.

**Theorem 3.** For an image  $I$  with size  $h \times w \times n_0$ , and an ASISR method constructed by  $\text{ASISR}_{\text{EQ}} = \Phi_{\text{INR}}^{\text{Glob}} [\Phi_{\text{CNN}}^{\text{EQ}}]$ , where  $\Phi_{\text{CNN}}^{\text{EQ}}$  is a Rot-E encoder constructed by  $L$ -layer rotation equivariant CNN network, whose channel number of the  $l^{\text{th}}$  layer is  $n_l$ , rotation equivariant subgroup is  $S \leq O(2)$ ,  $|S| = t$ , and activation function is set as ReLU.  $\Phi_{\text{INR}}^{\text{Glob}}$  denotes the global operator corresponding to the INR constructed with the proposed modules. Denote the latent continuous function of the  $c^{\text{th}}$  channel of  $I$  as  $r_c : \mathbb{R}^2 \rightarrow \mathbb{R}$ , and the latent continuous function of any convolution filters in the  $l^{\text{th}}$  layer as  $\phi^l : \mathbb{R}^2 \rightarrow \mathbb{R}$ . Assuming for any  $x \in \mathbb{R}^2$ ,  $l \in \{1, \dots, L\}$ ,  $c \in \{1, \dots, n_0\}$ , and any  $F$  produced with Rot-E encoder,  $A \in S$ , the following conditions are satisfied:

$$\begin{aligned} & \left\| \nabla_x \left( \Phi_{\text{INR}}^{\text{Glob}}(F)(x) \right)_c \right\| \leq G_{\text{INR}}^x, \\ & \sup_{i,j} \left\{ \left\| \nabla_{F_{ij}} \left( \Phi_{\text{INR}}^{\text{Glob}}(F)(x) \right)_c \right\| \right\} \leq G_{\text{INR}}^F, \\ & |r_c(x)| \leq F_0, \|\nabla_x r_c(x)\| \leq G_0, \|\nabla_x^2 r_c(x)\| \leq H_0, \\ & |\phi^l(x)| \leq F_l, \|\nabla_x \phi^l(x)\| \leq G_l, \|\nabla_x^2 \phi^l(x)\| \leq H_l, \\ & \forall \|x\| \geq (p+1)\delta/2, \varphi_l(x) = 0, \end{aligned} \quad (55)$$

where  $p$  is the filter size,  $\delta$  is the mesh size, and  $\nabla$  and  $\nabla^2$  denote the operators of gradient and Hessian matrix, respectively. Then, for any  $\tilde{A} \in S$  and  $x \in \mathbb{R}^2$ , the following result is satisfied:

$$|\text{ASISR}_{\text{EQ}} \left( \pi_{\tilde{A}}^I(I) \right)(x) - \pi_{\tilde{A}}^f [\text{ASISR}_{\text{EQ}}(I)](x)| \leq C\delta, \quad (56)$$

where  $\delta$  is the mesh size of the LR image,

$$C = \left( \sqrt{2n_L} G_{\text{INR}}^F \left( \sum_{m=1}^L \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F} + \sqrt{2} G_{\text{INR}}^x \right) + O(\delta), \quad (57)$$

and  $\mathcal{F} = \prod_{l=1}^L n_{l-1} p^2 F_l$ .

*Proof.* We can first deduce that

$$\begin{aligned} & |\text{ASISR}_{\text{EQ}} \left( \pi_{\tilde{A}}^I(I) \right)(x) - \pi_{\tilde{A}}^f [\text{ASISR}_{\text{EQ}}(I)](x)| \\ &= \left| \Phi_{\text{INR}}^{\text{Glob}} \left( \Phi_{\text{CNN}}^{\text{EQ}} \left[ \pi_{\tilde{A}}^I \right] (I) \right)(x) - \pi_{\tilde{A}}^f \left[ \Phi_{\text{INR}}^{\text{Glob}} \left[ \Phi_{\text{CNN}}^{\text{EQ}} \right] (I) \right](x) \right| \\ &\leq \left| \Phi_{\text{INR}}^{\text{Glob}} \left( \Phi_{\text{CNN}}^{\text{EQ}} \left[ \pi_{\tilde{A}}^I \right] (I) \right)(x) - \Phi_{\text{INR}}^{\text{Glob}} \left( \pi_{\tilde{A}}^F \left( \Phi_{\text{CNN}}^{\text{EQ}}(I) \right) \right)(x) \right| \\ &\quad + \left| \Phi_{\text{INR}}^{\text{Glob}} \left( \pi_{\tilde{A}}^F \left( \Phi_{\text{CNN}}^{\text{EQ}}(I) \right) \right)(x) - \pi_{\tilde{A}}^f \left[ \Phi_{\text{INR}}^{\text{Glob}} \left( \Phi_{\text{CNN}}^{\text{EQ}}(I) \right) \right](x) \right|. \end{aligned} \quad (58)$$

Since for  $\forall x$ ,  $\Phi_{\text{INR}}^{\text{Glob}}(F)(x) = \Phi_{\text{INR}}^{\text{Loc}}(F_{ij})(x - x_{ij})$ , where  $(i, j)$  is position in  $I$  whose coordinate is nearest to  $x$ , the value of  $\Phi_{\text{INR}}^{\text{Glob}}(F)(x)$  only correspond to 1 pixel in  $F$ . Therefore

$$\begin{aligned} & \left| \Phi_{\text{INR}}^{\text{Glob}} \left[ \Phi_{\text{CNN}}^{\text{EQ}} \left[ \pi_{\tilde{A}}^I \right] \right] (I)(x) - \Phi_{\text{INR}}^{\text{Glob}} \left( \pi_{\tilde{A}}^F \left( \Phi_{\text{CNN}}^{\text{EQ}}(I) \right) \right)(x) \right| \\ &\leq G_{\text{INR}}^F \sup_{i,j} \left\{ \left\| \left( \Phi_{\text{CNN}}^{\text{EQ}} \left[ \pi_{\tilde{A}}^I \right] (I) - \pi_{\tilde{A}}^F \left[ \Phi_{\text{CNN}}^{\text{EQ}} \right] (I) \right)_{ij} \right\| \right\} \\ &\leq 2\sqrt{n_L} G_{\text{INR}}^F \mathcal{F} \sum_{l=1}^L \sum_{m=1}^l \left( \frac{H_m F_0}{F_m} + 2 \sum_{n=1}^{m-1} \frac{G_m G_n F_0}{F_m F_n} + 2 \frac{G_m G_0}{F_m} + \frac{H_0}{l} \right) \delta^2, \end{aligned} \quad (59)$$

where we have applied Lemma 3 in the last line and  $\mathcal{F} = \prod_{l=1}^L n_{l-1} p^2 F_l$ . In the following we denote

$$C_1 = 2\sqrt{n_L} G_{\text{INR}}^F \mathcal{F} \sum_{l=1}^L \sum_{m=1}^l \left( \frac{H_m F_0}{F_m} + 2 \sum_{n=1}^{m-1} \frac{G_m G_n F_0}{F_m F_n} + 2 \frac{G_m G_0}{F_m} + \frac{H_0}{l} \right), \quad (60)$$

for convenience.

Besides, Let  $F = \Phi_{\text{CNN}}^{\text{EQ}}(I)$ , then is discredited from smooth function  $e^L(x, A)$  defined by (39), and by Lemma 2, we have

$$|\nabla_x e_c^L(x, B)| \leq \left( \sum_{m=1}^L \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}, \quad (61)$$

Then by using Theorem 2, we have

$$\begin{aligned} & \left| \Phi_{\text{INR}}^{\text{Glob}} \left( \pi_{\tilde{A}}^F \left( \Phi_{\text{CNN}}^{\text{EQ}}(I) \right) \right)(x) - \pi_{\tilde{A}}^f \left[ \Phi_{\text{INR}}^{\text{Glob}} \left( \Phi_{\text{CNN}}^{\text{EQ}}(I) \right) \right](x) \right| \\ & \leq \left( \sqrt{2n_L} G_{\text{INR}}^F \left( \sum_{m=1}^L \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F} + \sqrt{2} G_{\text{INR}}^x \right) \delta. \end{aligned} \quad (62)$$

Combining (58), (59) and (62), we have

$$\begin{aligned} & \left| \text{ASISR}_{\text{EQ}} \left( \pi_{\tilde{A}}^I(I) \right)(x) - \pi_{\tilde{A}}^f [\text{ASISR}_{\text{EQ}}(I)](x) \right| \\ & \leq \left( \sqrt{2n_L} G_{\text{INR}}^F \left( \sum_{m=1}^L \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F} + \sqrt{2} G_{\text{INR}}^x \right) \delta + C_1 \delta^2 \\ & = \left( \sqrt{2n_L} G_{\text{INR}}^F \left( \sum_{m=1}^L \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F} + \sqrt{2} G_{\text{INR}}^x + O(\delta) \right) \delta. \end{aligned} \quad (63)$$

This prove the conclusion.  $\square$

**Corollary 3.** Under the same conditions as Theorem 3, Moreover, when  $t = 2$  or  $t = 4$ , the following is satisfied for  $\forall \tilde{A} \in S$ :

$$\text{ASISR}_{\text{EQ}} \left( \pi_{\tilde{A}}^I(I) \right) = \pi_{\tilde{A}}^f [\text{ASISR}_{\text{EQ}}(I)]. \quad (64)$$

*Proof.* It is easy to find that when  $t = 2$  or  $t = 4$ , the rotation equivariance error of the utilized Rot-E encoder is zero [5]. Besides, with Corollary 2, we can see that the rotation equivariance error of the utilized Rot-E INR is also zero. Then with (58) we can see find that  $\forall x \in \mathbb{R}^2$ ,  $\text{ASISR}_{\text{EQ}} \left( \pi_{\tilde{A}}^I(I) \right)(x) = \pi_{\tilde{A}}^f [\text{ASISR}_{\text{EQ}}(I)](x)$ .  $\square$

## 7 PROOF OF COROLLARY 4

Before proving Corollary 4, we need to first provide the following two lemmas.

**Lemma 4** (Fu. 2023 [4]). For an image  $I$  with spatial size  $h \times w$  and mesh size  $\delta$ , denote the latent continuous function of  $I$  as  $r : \mathbb{R}^2 \rightarrow \mathbb{R}$ , and assume  $\|\nabla_x r(y)\| \leq G$ . Then, for arbitrary rotation degree, let  $\theta \in [0, 2\pi]$ ,  $A_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$  denotes the rotation matrix. If  $f(\theta) = r(A_\theta^{-1}(y))$ , then the following result is satisfied:

$$\left| \frac{\partial}{\partial \theta} f(\theta) \right| \leq (\max\{h, w\} + p) G \delta. \quad (65)$$

**Lemma 5.** For an image  $I$  with size  $h \times w \times n_0$ , and a ASISR method constructed by  $\text{ASISR}_{\text{EQ}} = \Phi_{\text{INR}}^{\text{Glob}} \left[ \Phi_{\text{CNN}}^{\text{EQ}} \right]$ , where  $\Phi_{\text{CNN}}^{\text{EQ}}$  is a Rot-E encoder constructed by  $L$ -layer rotation equivariant CNN network,  $\Phi_{\text{CNN}}^{\text{EQ}}(\cdot)$ , whose channel number of the  $l^{\text{th}}$  layer is  $n_l$ , rotation equivariant subgroup is  $S \leqslant O(2)$ ,  $|S| = t$ , and activation function is set as ReLU; and  $\Phi_{\text{INR}}^{\text{Glob}}$  denote the global operator corresponding to the INR constructed with the proposed modules, if the latent continuous function of the  $c^{\text{th}}$  channel of  $I$  denoted as  $r_c : \mathbb{R}^2 \rightarrow \mathbb{R}$ , and the latent continuous function of any convolution filters in the  $l^{\text{th}}$  layer denoted as  $\phi^l : \mathbb{R}^2 \rightarrow \mathbb{R}$ , for any  $x \in \mathbb{R}^2$ ,  $l \in \{1, \dots, L\}$ ,  $c \in \{1, \dots, n_0\}$ ,  $\forall F$  produce with Rot-E encoder,  $A \in S$ , the following conditions are satisfied:

$$\begin{aligned} & \sup_{i,j} \left\{ \left\| \nabla_{F_{ij}} \left( \Phi_{\text{INR}}^{\text{Glob}}(F)(x) \right)_c \right\| \right\} \leq G_{\text{INR}}^F, \\ & |r_c(x)| \leq F_0, \|\nabla_x r_c(x)\| \leq G_0, \|\nabla_x^2 r_c(x)\| \leq H_0, \\ & |\phi^l(x)| \leq F_l, \|\nabla_x \phi^l(x)\| \leq G_l, \|\nabla_x^2 \phi^l(x)\| \leq H_l, \\ & \forall \|x\| \geq (p+1)\delta/2, \phi_l(x) = 0, \end{aligned} \quad (66)$$

where  $p$  is the filter size,  $\delta$  is the mesh size, and  $\nabla$  and  $\nabla^2$  denote the operators of gradient and Hessian matrix, respectively. For an arbitrary  $\theta \in [0, 2\pi]$ ,  $A_\theta$  denotes the  $\theta$ -degree rotation matrix. Let  $f_x(\theta) = \text{ASISR}_{\text{EQ}} \left( \pi_{A_\theta}^I(I) \right)(x)$  and  $g_x(\theta) = \pi_{A_\theta}^f [\text{ASISR}_{\text{EQ}}(I)](x)$ , then the following result is satisfied  $\forall x \in \mathbb{R}^2$ :

$$\max \{|g'_x(\theta)|, |f'_x(\theta)|\} \leq n_L G_{\text{INR}}^F (\max\{h, w\} + Lp) \mathcal{F} G \delta, \quad (67)$$

where  $\mathcal{F} = \prod_{l=1}^L n_{l-1} F_l p^2$ .

*Proof.* In the following we will proof the conclusion without the ReLU activation functions for concise, while the proof are all still correct for networks with the ReLU activation function as ReLU does not disturb the equivariance or amplify the error bound.

Let  $F(\theta) = \Phi_{\text{CNN}}^{\text{EQ}} [\pi_{A_\theta}^I] (I)$ ,  $x$  be the coordinate of position  $(i, j)$ ,  $\Lambda$  be the local supporting set of filters (defined as (32)), then we can first deduce that

$$\begin{aligned} |(F(\theta))_{ij}^c| &= \left| (\Phi_L \star \cdots \star \Phi_l \star \cdots \star \Phi_2 \star \Psi \star (I))_{ij}^c \right| \\ &= \left| \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_1 \\ A_1 \in S \\ y_2 \in \Lambda}} \sum_{\substack{c_0 \\ y_1 \in \Lambda}} \phi_{A_{L-1} c_{L-1} c}^L (A_L^{-1} y_L) \cdots \phi_{A_1 c_1 c_2}^2 (A_2^{-1} y_2) \phi_{c_0 c_1}^1 (A_1^{-1} y_1) \nabla_\theta r_{c_0} (A_\theta^{-1} (x - y_L - \cdots - y_1)) \right| \\ &\leq \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_1 \\ A_1 \in S \\ y_2 \in \Lambda}} \sum_{\substack{c_0 \\ y_1 \in \Lambda}} \left| \phi_{A_{L-1} c_{L-1} c}^L (A_L^{-1} y_L) \right| \cdots \left| \phi_{A_1 c_1 c_2}^2 (A_2^{-1} y_2) \right| \left| \phi_{c_0 c_1}^1 (A_1^{-1} y_1) \right| \left| \nabla_\theta r_{c_0} (A_\theta^{-1} (x - y_L - \cdots - y_1)) \right| \\ &\leq \sum_{\substack{c_{L-1} \\ A_{L-1} \in S \\ y_L \in \Lambda}} \cdots \sum_{\substack{c_1 \\ A_1 \in S \\ y_2 \in \Lambda}} \sum_{\substack{c_0 \\ y_1 \in \Lambda}} F_L \cdots F_2 F_1 (\max\{h, w\} + L(p+1)) G\delta = (\max\{h, w\} + L(p+1)) \mathcal{F}G\delta \end{aligned} \quad (68)$$

Therefore,  $\forall c \in \{1, \dots, n_0\}$ ,

$$\begin{aligned} |(f_x(\theta))_c| &= |\nabla_\theta (\Phi_{\text{INR}}^{\text{Glob}}(F(\theta))(x))_c| \\ &\leq \sup_{c, i, j} \left\| \nabla_{F_{ij}} \left( \Phi_{\text{INR}}^{\text{Glob}}(F)(x) \right)_c \right\| \cdot \left\| (F(\theta))_{ij} \right\| \\ &\leq n_L G_{\text{INR}}^F (\max\{h, w\} + L(p+1)) \mathcal{F}G\delta. \end{aligned} \quad (69)$$

Similarly, we can prove that  $\|g_x(\theta)\| \leq n_L G_{\text{INR}}^F (\max\{h, w\} + L(p+1)) \mathcal{F}G\delta$ , this complete the proof.  $\square$

**Corollary 4.** Under the same conditions as Theorem 3, for an arbitrary  $\theta \in [0, 2\pi]$ , let  $A_\theta$  denotes the  $\theta$ -degree rotation matrix, then  $\forall \theta$  we have

$$\left| \text{ASISR}_{\text{EQ}} \left( \pi_{A_\theta}^I (I) \right) (x) - \pi_{A_\theta}^f [\text{ASISR}_{\text{EQ}} (I)] (x) \right| \leq C\delta + \hat{C}\delta t^{-1}, \quad (70)$$

where  $\hat{C} = 4\pi n_L G_{\text{INR}}^F (\max\{h, w\} + L(p+1))$ .

*Proof.* It is easy to find that  $S$  is indeed a group of rotation matrixes whose rotation degree are in the set  $\Theta = \{\theta_k = \frac{2\pi k}{t} | k = 1, 2, \dots, t\}$ . Then, for  $\forall \theta$ , let  $f_x(\theta) = \text{ASISR}_{\text{EQ}} (\pi_{A_\theta}^I (I)) (x)$  and  $g_x(\theta) = \pi_{A_\theta}^I [\text{ASISR}_{\text{EQ}} (I)] (x)$ , and let  $\theta^* = \min_{\hat{\theta} \in \Theta} |\theta - \hat{\theta}|$ , we can find  $A_{\theta^*} \in S$  and  $|\theta^* - \theta| \leq \frac{2\pi}{t}$ . With Theorem 3 and Lemma 5, it is easy to deduce that

$$\begin{aligned} &\left| \text{ASISR}_{\text{EQ}} \left( \pi_{A_\theta}^I (I) \right) (x) - \pi_{A_\theta}^f [\text{ASISR}_{\text{EQ}} (I)] (x) \right| \\ &\leq \left| \text{ASISR}_{\text{EQ}} \left( \pi_{A_\theta}^I (I) \right) (x) - \text{ASISR}_{\text{EQ}} \left( \pi_{A_{\theta^*}}^I (I) \right) (x) \right| + \left| \text{ASISR}_{\text{EQ}} \left( \pi_{A_{\theta^*}}^I (I) \right) (x) - \pi_{A_{\theta^*}}^f [\text{ASISR}_{\text{EQ}} (I)] (x) \right| + \\ &\quad \left| \pi_{A_{\theta^*}}^f [\text{ASISR}_{\text{EQ}} (I)] (x) - \pi_{A_\theta}^f [\text{ASISR}_{\text{EQ}} (I)] (x) \right| \\ &\leq \sup_{c, \theta} \{|(f'_x(\theta))_c|\} |\theta^* - \theta| + C_1 \delta^2 + C_2 + \delta + \sup_{c, \theta} \{|(g'_x(\theta))_c|\} |\theta^* - \theta| \\ &\leq C_1 \delta^2 + C_2 \delta + 4\pi n_L G_{\text{INR}}^F (\max\{h, w\} + Lp) \mathcal{F}G\delta t^{-1}, \end{aligned} \quad (71)$$

This complete the proof.  $\square$

## 8 PROOF OF REMARK 1

**Remark 1.** Set  $t = 4$ , then, by replacing the linear patch embedding layer with Rot-E convolution layer, replacing all the linear layer with the Rot-E linear layer (i.e., Rot-E convolution layer with kernel size of  $1 \times 1$ ), and removing the position embedding module, the simplified Vision-Transformer-based encoder and Swin-Transformer encoder are  $\frac{\pi}{2}$  rotation equivariant.

*Proof.* 1) For Vision-Transformer-based encoder, there are linear embedding module (linear projection of flattened patches), position embedding layers and transformer modules. Since in the utilized simplified Vision-Transformer-encoder, position embedding layers have been removed and the linear embedding module have been replaces with Rot-E convolution layer ( $t = 4$ ), we only need to prove the Transformer constructed with Rot-E linear layer is  $\frac{\pi}{2}$  rotation equivariant.

Formally, we can write a Transformer layer as

$$\text{Transformer}(F) = \text{fold} \left( \text{softmax} \left( \frac{\text{unfold}(\mathcal{L}_K(F)) \cdot \text{unfold}(\mathcal{L}_Q(F))^T}{d_k} \right) \cdot \text{unfold}(\mathcal{L}_V(F)) \right), \quad (72)$$

where  $F$  provide by the input Rot-E convolution layer or the previous Transformer layer,  $\hat{F}$  is the output layer,  $d_k$  is a selected constant,  $\mathcal{L}_K$ ,  $\mathcal{L}_Q$  and  $\mathcal{L}_V$  are Rot-E linear layers;  $\text{unfold}(\cdot)$  defines the unfolding operation that reshape a 4d tensor in the shape  $(h \times w \times n \times t)$  into a matrix of shape  $(hw \times nt)$  and  $\text{fold}(\cdot)$  is its inverse operation.

For  $\forall a \leq hw$  and  $p \leq nt$ , let their corresponding pixel index and channel index be  $(i_a, j_a)$  and  $(k_p, l_p)$ , respectively. Then,  $\forall F$  we have

$$(\text{unfold}(F))_{ap} = F_{i_a j_a}^{k_p l_p}. \quad (73)$$

Further more, for arbitrary rotation equivariant feature maps  $\bar{F}$ ,  $\hat{F}$  and  $\tilde{F}$ , we have

$$\begin{aligned} & \left( \text{fold} \left( \text{softmax} \left( \frac{\text{unfold}(\bar{F}) \cdot \text{unfold}(\hat{F})^T}{d_k} \right) \cdot \text{unfold}(\tilde{F}) \right) \right)_{i_a j_a}^{k_p l_p} \\ &= \left( \text{softmax} \left( d_k^{-1} \text{unfold}(\bar{F}) \cdot \text{unfold}(\hat{F})^T \right) \cdot \text{unfold}(\tilde{F}) \right)_{ap} \\ &= \sum_b \left( \text{softmax} \left( d_k^{-1} \text{unfold}(\bar{F}) \cdot \text{unfold}(\hat{F})^T \right) \right)_{ab} \left( \text{unfold}(\tilde{F}) \right)_{bp} \\ &= \frac{\sum_b \exp \left( d_k^{-1} \left( \text{unfold}(\bar{F}) \cdot \text{unfold}(\hat{F})^T \right)_{ab} \right) \tilde{F}_{i_b j_b}^{k_p l_p}}{\sum_b \exp \left( d_k^{-1} \left( \text{unfold}(\bar{F}) \cdot \text{unfold}(\hat{F})^T \right)_{ab} \right)} \\ &= \frac{\sum_b \exp \left( d_k^{-1} \sum_q (\text{unfold}(\bar{F}))_{aq} \left( \text{unfold}(\hat{F}) \right)_{bq} \right) \tilde{F}_{i_b j_b}^{k_p l_p}}{\sum_b \exp \left( d_k^{-1} \sum_q (\text{unfold}(\bar{F}))_{aq} \left( \text{unfold}(\hat{F}) \right)_{bq} \right)} \\ &= \frac{\sum_b \exp \left( d_k^{-1} \sum_q \bar{F}_{i_a j_a}^{k_q l_q} \hat{F}_{i_b j_b}^{k_q l_q} \right) \tilde{F}_{i_b j_b}^{k_p l_p}}{\sum_b \exp \left( d_k^{-1} \sum_q \bar{F}_{i_a j_a}^{k_q l_q} \hat{F}_{i_b j_b}^{k_q l_q} \right)}. \end{aligned} \quad (74)$$

For any spatial index  $(i, j)$ , group index  $l$ , and  $\frac{k\pi}{2}$  rotation  $k = 0, 1, 2, 3$ , let  $\tilde{A}$  be the rotation matrix of  $\frac{k\pi}{2}$  degree,  $(\tilde{i}, \tilde{j})$  define the index of rotation result of pixel  $x_{ij}$ , i.e.,  $x_{\tilde{i}\tilde{j}} = \tilde{A}x_{ij}$ ,  $\tilde{l}$  define the  $k$  cyclically shifting result of channel index  $l$ , then,  $\forall F$  we have

$$(\pi_{\tilde{A}}^F(F))_{i_a j_a}^{k_p l_p} = F_{\tilde{i}_a \tilde{j}_a}^{k_p \tilde{l}_p}. \quad (75)$$

Then, we can deduce that

$$\begin{aligned} & \left( \text{fold} \left( \text{softmax} \left( \frac{\text{unfold}(\mathcal{L}_K(\pi_{\tilde{A}}^F(F))) \cdot \text{unfold}(\mathcal{L}_Q(\pi_{\tilde{A}}^F(F))^T}{d_k} \right) \cdot \text{unfold}(\mathcal{L}_V(\pi_{\tilde{A}}^F(F))) \right) \right)_{i_a j_a}^{k_p l_p} \\ &= \left( \text{fold} \left( \text{softmax} \left( \frac{\text{unfold}(\pi_{\tilde{A}}^F(\mathcal{L}_K(F))) \cdot \text{unfold}(\pi_{\tilde{A}}^F(\mathcal{L}_Q(F))^T}{d_k} \right) \cdot \text{unfold}(\pi_{\tilde{A}}^F(\mathcal{L}_V(F))) \right) \right)_{i_a j_a}^{k_p l_p} \\ &= \frac{\sum_b \exp \left( d_k^{-1} \sum_q \left( \pi_{\tilde{A}}^F(\mathcal{L}_K(F)) \right)_{i_a j_a}^{k_q l_q} \left( \pi_{\tilde{A}}^F(\mathcal{L}_Q(F)) \right)_{i_b j_b}^{k_q l_q} \right) \left( \pi_{\tilde{A}}^F(\mathcal{L}_V(F)) \right)_{i_b j_b}^{k_p l_p}}{\sum_b \exp \left( d_k^{-1} \sum_q \left( \pi_{\tilde{A}}^F(\mathcal{L}_K(F)) \right)_{i_a j_a}^{k_q l_q} \left( \pi_{\tilde{A}}^F(\mathcal{L}_Q(F)) \right)_{i_b j_b}^{k_q l_q} \right)} \\ &= \frac{\sum_b \exp \left( d_k^{-1} \sum_q (\mathcal{L}_K(F))_{\tilde{i}_a \tilde{j}_a}^{k_q \tilde{l}_q} (\mathcal{L}_Q(F))_{\tilde{i}_a \tilde{j}_a}^{k_q \tilde{l}_q} \right) (\mathcal{L}_V(F))_{\tilde{i}_a \tilde{j}_a}^{k_q \tilde{l}_q}}{\sum_b \exp \left( d_k^{-1} \sum_q (\mathcal{L}_K(F))_{\tilde{i}_a \tilde{j}_a}^{k_q \tilde{l}_q} (\mathcal{L}_Q(F))_{\tilde{i}_a \tilde{j}_a}^{k_q \tilde{l}_q} \right)} \\ &= \left( \text{fold} \left( \text{softmax} \left( \frac{\text{unfold}(\mathcal{L}_K(F)) \cdot \text{unfold}(\mathcal{L}_Q(F))^T}{d_k} \right) \cdot \text{unfold}(\mathcal{L}_V(F)) \right) \right)_{\tilde{i}_a \tilde{j}_a}^{k_q \tilde{l}_q} \\ &= \left( \pi_{\tilde{A}}^F \left( \text{fold} \left( \text{softmax} \left( \frac{\text{unfold}(\mathcal{L}_K(F)) \cdot \text{unfold}(\mathcal{L}_Q(F))^T}{d_k} \right) \cdot \text{unfold}(\mathcal{L}_V(F)) \right) \right) \right)_{i_a j_a}^{k_p l_p}. \end{aligned} \quad (76)$$

In other words, we have

$$\text{Transformer}(\pi_{\tilde{A}}^F(F)) = \pi_{\tilde{A}}^F(\text{Transformer}(F)), \quad (77)$$

which prove that the  $\frac{\pi}{2}$  rotation equivariance of the defined Transformer.

2) Shift-Window-Transformer-encoder is composed of shifted window operation and Transformer modules. Clearly, the shifted window operation is  $\frac{\pi}{2}$  rotation equivariant, and we have proved that the Transformer module constructed in the aforementioned manner is  $\frac{\pi}{2}$  rotation equivariant. Thus, the Swin-Transformer-encoder construction in the proposed manner is  $\frac{\pi}{2}$  rotation equivariant.  $\square$

## 9 MORE EQUIVARIANCE VERIFICATION EXPERIMENTS

In the main text, we have visually illustrated the equivariance errors of LTE and its Rot-E improvement, i.e., LTE-EQ (in Fig. 8 of the main text). However, due to the page limitation, we have not included corresponding illustrations for LIIF and OPE in the main text.

We thus want to demonstrate more visualization of equivariance errors in this section. Fig. 2 illustrates rotation equivariance error maps for the three ASISR methods (LIIF, OPE and LTE), at  $90^\circ$ ,  $180^\circ$  and  $270^\circ$  rotations, evaluated on randomly initialized, untrained networks. Besides, Fig. 3 subsequently shows these error maps for all the six models after 100 training epochs.

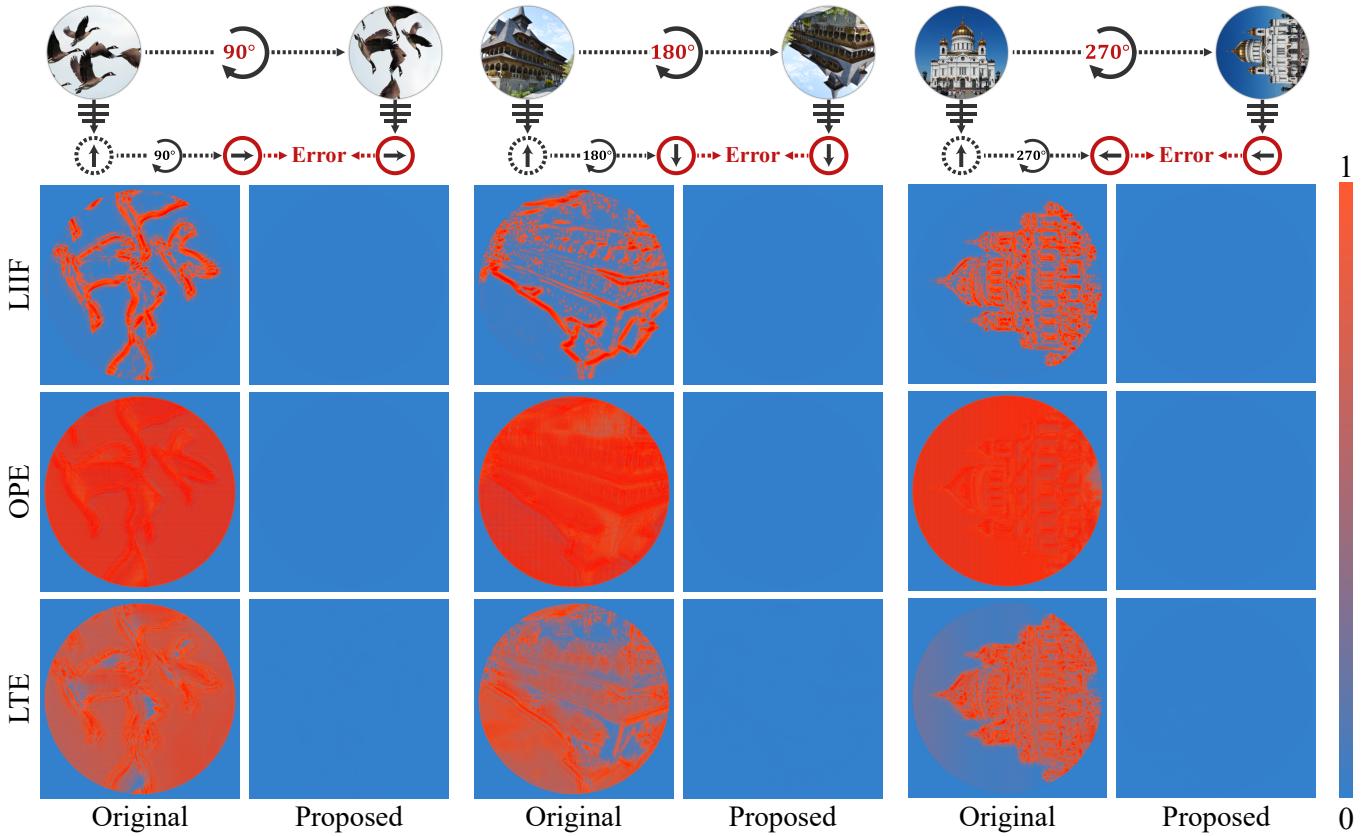


Fig. 2. Illustration of rotation equivariance errors ( $|\Phi\pi_{\hat{A}}^I - \pi_{\hat{A}}^F\Phi|$  for image  $I$  and network  $\Phi$ ) comparing original LIIF, OPE, LTE with their rotation-equivariant improvements (p16 rotation equivariant). Evaluations are performed at rotation angles of  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ , with all networks randomly initialized without any training.

From these two figures, it can be easily observed that when being rotated for  $90^\circ$ ,  $180^\circ$  and  $270^\circ$ , traditional ASISR methods exhibit evident rotation equivariance errors, regardless of whether the network has been trained or not. In contrast, our proposed improvements demonstrate significantly smaller rotation equivariance errors (nearly zero), confirming that the framework effectively enhances the rotation equivariance of ASISR methods. Moreover, this result is finely consistent with our theoretical result in Corollary 3, which also verifies the correctness of our theoretical findings.

## 10 MORE GENERALIZATION RESULTS ON HYPERSPECTRAL IMAGE SUPER-RESOLUTION

To validate the generalization ability of the proposed method, we applied the models trained in the DIV2K natural image dataset [6] to the super-resolution tasks in the CAVE hyperspectral image (HSI) dataset [7].

In the main text, we compare the proposed method with competing methods on the HSIs in the CAVE dataset. Specifically, we process SR for each band of the HSI one by one. Since the images in the DIV2K dataset are color images, the model trained on this dataset cannot directly process each band of the HSI (which has only one channel). Therefore,

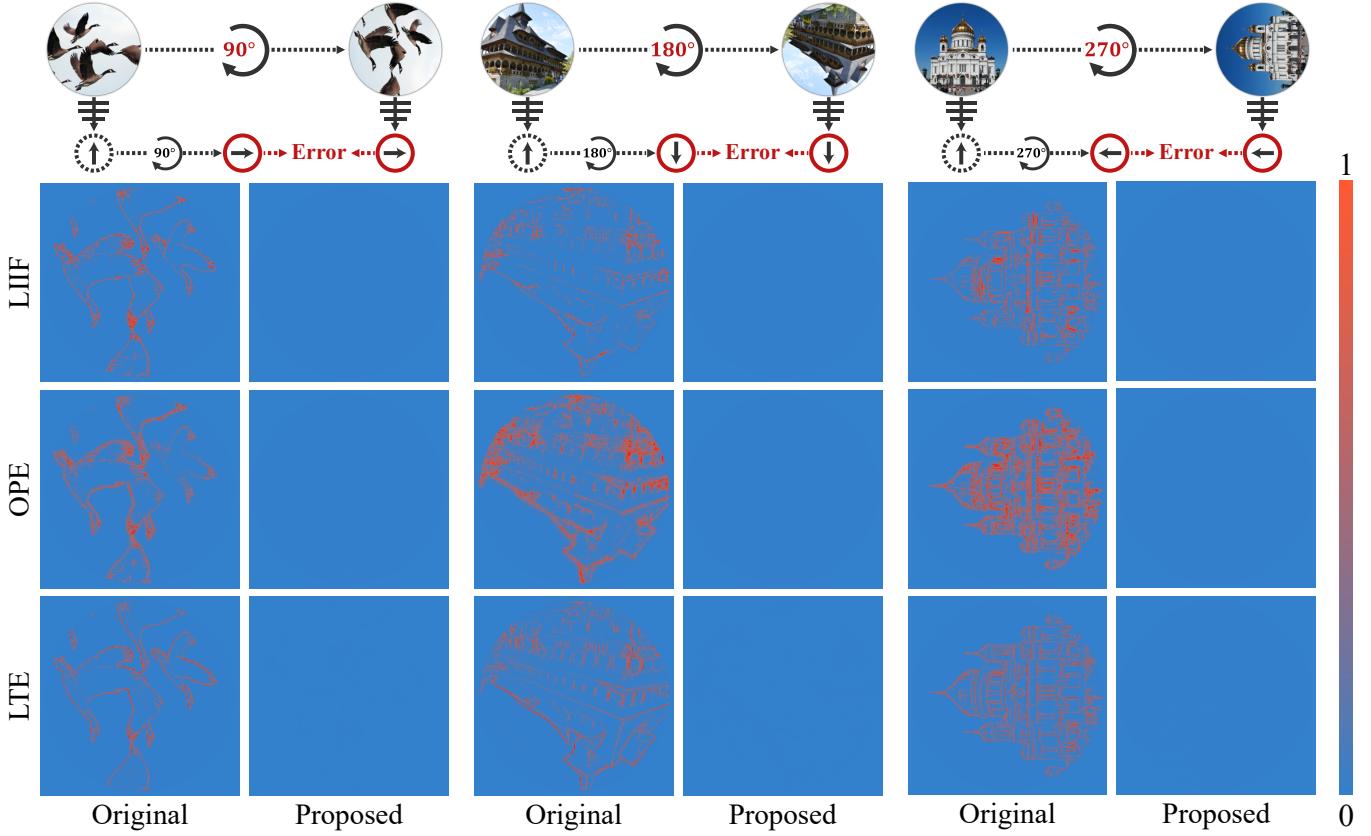


Fig. 3. Illustration of rotation equivariance errors ( $|\Phi\pi_A^I - \pi_A^F\Phi|$  for image  $I$  and network  $\Phi$ ) comparing original LIIF, OPE, LTE with their rotation-equivariant improvements (p16 rotation equivariant). Evaluations are performed at rotation angles of  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ , with all networks trained for 100 epochs.

we duplicated each band of the HSI three times and treated it as an RGB image to perform ASISR. The test results can be found in Table 5 of the main text.

It should be noted that there is a significant difference in image types between DIV2K and CAVE, as shown in Fig. 4, where the hyperspectral data in CAVE is visualized by selecting the 24th, 12th and 8th bands to compose the RGB channels. We can see that the DIV2K dataset consists mainly of outdoor images, whereas the CAVE dataset primarily consists of indoor images.

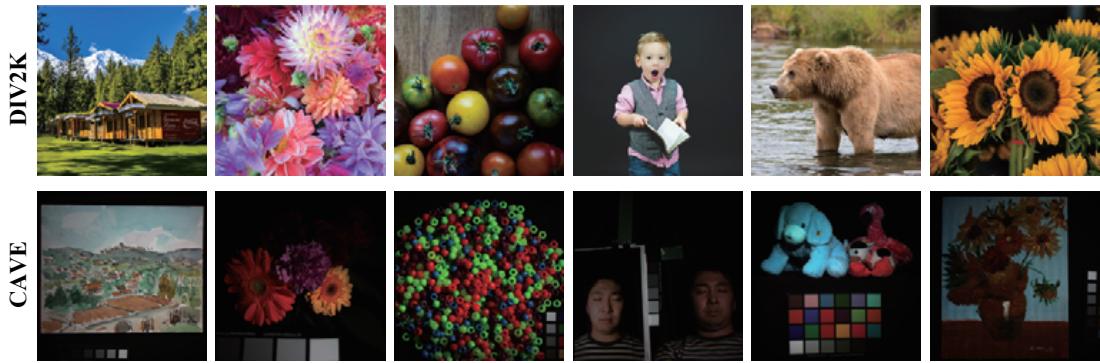


Fig. 4. Visualization comparison of sample images from the CAVE [7] and DIV2K [6] Datasets.

Therefore, we believe that, beyond testing the performance on HSIs, it is also valuable to test the generalization ability of the proposed methods on RGB images from the CAVE dataset. Specifically, we test the performance on all the three aforementioned ASISR methods (i.e., LIIF [8], OPE [9] and LTE [10]), and consistently set their encoder as the aforementioned EDSR-baseline encoder [11]. As shown in Table 1, compared to the original method, the proposed method further improves the reconstruction performance of RGB images in the CAVE data set.

TABLE 1

The generalization results on RGB images in the CAVE Dataset [7] are presented. All the competing models are trained on the DIV2K dataset [6], and the PSNR (dB) results are averaged over all the samples in the CAVE dataset.

Method	Mode	x2	x3	x4	x6	x8	x12
LIIF [8]	Original	40.74	38.02	36.27	33.88	32.09	29.52
	Proposed	<b>40.84</b>	<b>38.15</b>	<b>36.38</b>	<b>34.00</b>	<b>32.19</b>	<b>29.61</b>
OPE [9]	Original	40.69	37.92	36.19	33.75	31.95	<b>29.37</b>
	Proposed	<b>40.70</b>	<b>37.97</b>	<b>36.22</b>	<b>33.79</b>	<b>31.98</b>	29.36
LTE [10]	Original	40.74	38.04	36.28	33.94	32.11	29.51
	Proposed	<b>40.83</b>	<b>38.14</b>	<b>36.37</b>	<b>33.97</b>	<b>32.24</b>	<b>29.59</b>

## 11 FAILURE CASES ANALYSIS

In a large number of experiments, we also have found certain cases where the proposed methods yield unsuccessful reconstruction results. In this section, we present a typical failure example, which would help us better understand the limitations of our method and gain insights into potential areas for further improvement.

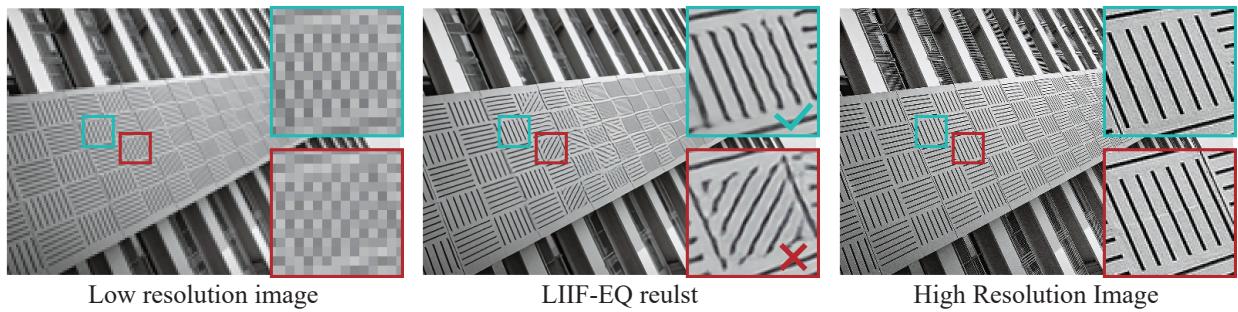


Fig. 5. Illustration of a typical instance where errors occur in the direction of the reconstructed texture when the super-resolution scale is 6.

As shown Fig. 5, when the resolution of the input image is very low, the proposed method might reconstruct textures with incorrect directions. As shown in the figure, for input textures in the same direction, different reconstruction results may be obtained. This is a typical case of failure to maintain equivariance.

This problem could be rationally attributed to the severe degradation of input textures in low-resolution images, making it extremely difficult to reconstruct its correct directional information through the local feature. According to the proposed theory (Corollary 4), the equivariance error is determined by the local derivative values ( $G_0$  and  $H_0$ ) and the mesh size ( $\delta$ ) of the input image. When the texture of the input image is complex and the resolution is too low, both the local derivative values and the mesh size are large. Consequently, the equivariance error will also increase, which may lead to incorrect reconstruction results.

To alleviate this issue, it may be necessary to increase the utilization of non-local correlations, which have been comprehensively substantiated as a strong prior knowledge in image restoration tasks [12], [13], in different regions during the design of the INR. By leveraging the information complementarity of different regions, the errors could be hopeful to be reduced in the direction of the reconstructed texture in regions with extremely low resolution. This could be a future research direction worthy of exploration.

## REFERENCES

- [1] Zhengyang Shen, Lingshen He, Zhouchen Lin, and Jinwen Ma. Pdo-econvs: Partial differential operator based equivariant convolutions. In *International Conference on Machine Learning*, pages 8697–8706. PMLR, 2020.
- [2] Maurice Weiler and Gabriele Cesa. General e (2)-equivariant steerable cnns. *Advances in Neural Information Processing Systems*, 32, 2019.
- [3] Qi Xie, Qian Zhao, Zongben Xu, and Deyu Meng. Fourier series expansion based filter parametrization for equivariant convolutions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.
- [4] Jiahong Fu, Qi Xie, Deyu Meng, and Zongben Xu. Rotation equivariant proximal operator for deep unfolding methods in image restoration. *arXiv preprint arXiv:2312.15701*, 2023.
- [5] Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pages 2990–2999. PMLR, 2016.
- [6] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017.
- [7] Fumihito Yasuma, Tomoo Mitsunaga, Daisuke Iso, and Shree K Nayar. Generalized assorted pixel camera: postcapture control of resolution, dynamic range, and spectrum. *IEEE transactions on image processing*, 19(9):2241–2253, 2010.
- [8] Yinbo Chen, Sifei Liu, and Xiaolong Wang. Learning continuous image representation with local implicit image function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8628–8638, 2021.
- [9] Gaochao Song, Qian Sun, Luo Zhang, Ran Su, Jianfeng Shi, and Ying He. Ope-sr: Orthogonal position encoding for designing a parameter-free upsampling module in arbitrary-scale image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10009–10020, 2023.

- [10] Jaewon Lee and Kyong Hwan Jin. Local texture estimator for implicit representation function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1929–1938, 2022.
- [11] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.
- [12] Shuhang Gu, Qi Xie, Deyu Meng, Wangmeng Zuo, Xiangchu Feng, and Lei Zhang. Weighted nuclear norm minimization and its applications to low level vision. *International journal of computer vision*, 121:183–208, 2017.
- [13] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7794–7803, 2018.