

Supplementary Material for CVPR'2025 Paper: Rotation-Equivariant Self-Supervised Method in Image Denoising

Abstract

This supplementary material presents more details on the theoretical results in the main paper. More implementation details of the proposed method and more experimental results are also provided for better reference to the readers. The structure of the supplementary material is organized as follows. In the first section, we introduce the necessary notations and definitions, laying the foundation for the subsequent analysis. Building on this, the second section presents the rotation equivariant errors of two commonly used downsampling operators, accompanied by detailed proof processes to ensure clarity and rigor. Similarly, the third section focuses on the rotation equivariant errors of two common upsampling operators, again providing comprehensive proofs to substantiate the findings. In the fourth section, we take the U-Net network from the N2N method as an illustrative example. Within this context, we derive and present the equivariant error bounds for the complete U-Net network under fixed discrete angles, formulated as theorems. Extending this analysis, we further provide the equivariant error bounds for the complete U-Net network under arbitrary rotation angles, which are summarized in the form of corollaries. In the fifth section, we present additional experimental results and visualizations to further demonstrate the effectiveness of the proposed method.

I. NOTATIONS AND DEFINITIONS

We first introduce some necessary notations and preliminaries as follows.

We consider the equivariance on the orthogonal group $O(2)$. Formally, $O(2) = \{A \in \mathbb{R}^{2 \times 2} | A^T A = I_{2 \times 2}\}$, which contains all rotation and reflection matrices. Without ambiguity, we use A to parameterize $O(2)$. The Euclidean group $E(2) = \mathbb{R}^2 \rtimes O(2)$ (\rtimes is a semidirect-product), whose element is represented as (x, A) . Restricting the domain of A and x , we can also use this representation to parameterize any subgroup of $E(2)$. In practice, the subgroup is usually assumed to contain t rotations with $\frac{2\pi}{t}$ degree for an integer $t \in \mathbb{N}_+$.

An image $I \in \mathbb{R}^{n \times n}$ is viewed as a two-dimensional discretization of a smooth function $r : \mathbb{R}^2 \rightarrow \mathbb{R}$, at the cell-center of a regular grid with $n \times n$ cells, i.e., for $i, j = 1, 2, \dots, n$,

$$I_{ij} = r(x_{ij}), \quad (1)$$

where $x_{ij} = \left(\left(i - \frac{n+1}{2} \right) h, \left(j - \frac{n+1}{2} \right) h \right)^T$ and h is the mesh size.

An intermediate feature map $F \in \mathbb{R}^{n \times n \times t}$ in equivariant networks is a multi-channel tensor, which can be viewed as the discretization of a continuous function defined on $\tilde{E} = \mathbb{R}^2 \rtimes S$, where S is a subgroup of $O(2)$ and t is the number of elements in S . Formally, F can be represented as a three-dimensional grid tensor sampled from a smooth function $e : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$, i.e., for $i, j = 1, 2, \dots, n$,

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

where x_{ij} is defined in (1) and $A \in S$.

With above notations, the transformations on the input and feature maps can be mathematically formulated. Specifically, in the continuous domain, for an input $r \in C^\infty(\mathbb{R}^2)$ and feature map $e \in C^\infty(E(2))$, the transformation $\tilde{A} \in O(2)$ acts on r and e respectively by:

$$\begin{aligned} \pi_{\tilde{A}}^R[r](x) &= r(\tilde{A}^{-1}x), \forall x \in \mathbb{R}^2, \\ \pi_{\tilde{A}}^E[e](x, A) &= e(\tilde{A}^{-1}x, \tilde{A}^{-1}A), \forall (x, A) \in E(2). \end{aligned} \quad (3)$$

In particular, if $A_\theta \in O(2)$ is the rotation matrix $\begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$, then the corresponding rotation operators can be expressed by π_θ^R and π_θ^E .

Besides, in the discrete domain, we can also define the transformation $\tilde{A} \in S$ on the input image and feature map as followings:

$$\begin{aligned} (\tilde{\pi}_{\tilde{A}}^R(I))_{ij} &= \pi_{\tilde{A}}^R[r](x_{ij}), \\ (\tilde{\pi}_{\tilde{A}}^E(F))_{ij}^A &= \pi_{\tilde{A}}^E[e](x_{ij}, A), \\ \forall i, j &= 1, 2, \dots, n, A \in S. \end{aligned} \quad (4)$$

Similarly, the rotation operators can be denoted as $\tilde{\pi}_\theta^R$ and $\tilde{\pi}_\theta^E$.

Before the proof of the theorem, we provide the definitions of commonly used downsampling and upsampling methods in the continuous domain.

Maxpooling Downsampling. Maxpooling is a commonly used downsampling method in CNNs, which reduces the spatial dimensions of feature maps by sliding a fixed-size window over the feature map and selecting the maximum value within each region as the output [1]. In the continuous domain, we can define maxpooling operator $M(\cdot)$ as follows,

$$[M(F)](x, A) = \max\{F_{ij}^A, F_{i+1,j}^A, F_{i,j+1}^A, F_{i+1,j+1}^A\} = \max_{\Omega_{ij}} F_{ij}^A, \quad (5)$$

where $x = [x_1, x_2]^T \in \mathbb{R}^2$ denotes the spatial coordinates, and $x_1 \in [x_{ij_1}, x_{i+1,j_1}]$, $x_2 \in [x_{ij_2}, x_{i,j+1_2}]$, $\Omega_{ij} = \{(i, j), (i+1, j), (i, j+1), (i+1, j+1)\}$.

Stride Downsampling. Stride Downsampling is also a widely used downsampling operator which reduce the size of the feature map by adjusting the stride of the convolution operation [2]. In the continuous domain, we can define stride downsampling operator $S(\cdot)$ as follows,

$$[S(F)](x, A) = F_{i,j+1}^A, \quad (6)$$

where $x = [x_1, x_2]^T \in \mathbb{R}^2$ denotes the spatial coordinates, and $x_1 \in [x_{ij_1}, x_{i+1,j_1}]$, $x_2 \in [x_{ij_2}, x_{i,j+1_2}]$.

Nearest Neighbor Upsampling. Nearest neighbor interpolation is an image scaling method that fills the pixels of the interpolated image by selecting the original pixel value closest to the target pixel position. In the continuous domain, we can define the nearest neighbor operator $N(\cdot)$ as follows,

$$[N(F)](x, A) = F_{i^*j^*}^A, \quad (7)$$

where $(i^*, j^*) = \arg \min_{ij} \|x_{ij} - x\|_2^2$.

Bilinear Upsampling. Bilinear interpolation calculates the new pixel value by taking the weighted average of the four surrounding known pixel values. In the continuous domain, we can define the bilinear interpolation operator $B(\cdot)$ as follows,

$$\begin{aligned} [B(F)](x, A) &= \frac{(v_2 - x_2)}{(v_2 - v_1)} \left[\frac{(u_2 - x_1)}{(u_2 - u_1)} f(Q_{11}) + \frac{(x_1 - u_1)}{(u_2 - u_1)} f(Q_{21}) \right] + \frac{(x_2 - v_1)}{(v_2 - v_1)} \left[\frac{(u_2 - x_1)}{(u_2 - u_1)} f(Q_{12}) + \frac{(x_1 - u_1)}{(u_2 - u_1)} f(Q_{22}) \right] \\ &= \lambda_{11} f(Q_{11}) + \lambda_{21} f(Q_{21}) + \lambda_{12} f(Q_{12}) + \lambda_{22} f(Q_{22}) \\ &= \sum_{i=1}^2 \sum_{j=1}^2 \lambda_{ij} f(Q_{ij}), \end{aligned} \quad (8)$$

where λ_{ij} are the coefficients of bilinear interpolation and $f(Q_{ij})$ represent the grid points, $x = [x_1, x_2]^T \in \mathbb{R}^2$ denotes the 2D spatial coordinates, $x_1 \in [x_{ij_1}, x_{i+1,j_1}]$, $x_2 \in [x_{ij_2}, x_{i,j+1_2}]$.

II. PROOF OF EQUIVARIANCE ERROR FOR DOWNSAMPLING OPERATORS

A. Proof of equivariance error for maxpooling downsampling

Theorem 1. Assume that a feature map $F \in \mathbb{R}^{n \times n \times t}$ is discretized from the smooth function $e : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$, $|S|=t$, the mesh size is h , $M(\cdot)$ is the downsampling operator. If for any $A, B \in S, x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\| \nabla e(x, A) \| \leq G, \quad (9)$$

then the following results are satisfied:

$$|M[\tilde{\pi}_B^E](F)(x, A) - \pi_B^E[M(F)](x, A)| \leq 2\sqrt{2}Gh. \quad (10)$$

Proof. From the above definition we can deduce as follows:

$$\begin{aligned} [M[\tilde{\pi}_B^E](F)](x, A) &= M(\tilde{\pi}_B^E(F))(x, A) = \max\{(\tilde{\pi}_B^E(F))_{ij}^A, (\tilde{\pi}_B^E(F))_{i+1,j}^A, (\tilde{\pi}_B^E(F))_{i,j+1}^A, (\tilde{\pi}_B^E(F))_{i+1,j+1}^A\} \\ &= \max\{e(B^{-1}x_{ij}, B^{-1}A), e(B^{-1}x_{i+1,j}, B^{-1}A), e(B^{-1}x_{i,j+1}, B^{-1}A), e(B^{-1}x_{i+1,j+1}, B^{-1}A)\} \\ &= \max_{\Omega_{ij}} e(B^{-1}x_{ij}, B^{-1}A), \end{aligned} \quad (11)$$

$$\begin{aligned} \pi_B^E[M(F)](x, A) &= M(F)(B^{-1}x, B^{-1}A) \\ &= \max\{e(x_{i'j'}, B^{-1}A), e(x_{i'+1,j'}, B^{-1}A), e(x_{i',j'+1}, B^{-1}A), e(x_{i'+1,j'+1}, B^{-1}A)\} \\ &= \max_{\Omega_{i'j'}} e(x_{i'j'}, B^{-1}A), \end{aligned} \quad (12)$$

where $x = [x_1, x_2]^T \in \mathbb{R}^2$, $B^{-1}x = [x'_1, x'_2]^T \in \mathbb{R}^2$ denotes the spatial coordinates, and $x_1 \in [x_{ij_1}, x_{i+1, j_1}]$, $x_2 \in [x_{ij_2}, x_{i, j_2+1}]$, $x'_1 \in [x_{i'j'_1}, x_{i'+1, j'_1}]$, $x'_2 \in [x_{i'j'_2}, x_{i', j'_2+1}]$. We define $x_{i\hat{j}}$ as the coordinates of the maximum point in (11), $x_{i\hat{j}}$ as the coordinates of the maximum point in (12). By the definition of $x_{i\hat{j}}$, we have $\|x_{i\hat{j}} - x\|_2^2 \leq \sqrt{2}h$. Since B is an orthogonal matrix, we have

$$\|B^{-1}x_{i\hat{j}} - B^{-1}x\|_2^2 \leq \sqrt{2}h. \quad (13)$$

By the definition of $x_{i\hat{j}}$,

$$\|x_{i\hat{j}} - B^{-1}x\|_2^2 \leq \sqrt{2}h. \quad (14)$$

From (13) and (14) we can derive

$$\|B^{-1}x_{i\hat{j}} - x_{i\hat{j}}\|_2^2 \leq \|B^{-1}x_{i\hat{j}} - B^{-1}x\|_2^2 + \|x_{i\hat{j}} - B^{-1}x\|_2^2 \leq \sqrt{2}h + \sqrt{2}h = 2\sqrt{2}h. \quad (15)$$

From (9) and (15) and by using the Lagrange Mean Value Theorem, we can derive:

$$|M[\tilde{\pi}_B^E](F)(x, A) - \pi_B^E[M(F)](x, A)| = |e(B^{-1}x_{i\hat{j}}, B^{-1}A) - e(x_{i\hat{j}}, B^{-1}A)| \leq 2\sqrt{2}Gh. \quad (16)$$

□

B. Proof of equivariance error for stride downsampling

Theorem 2. Assume that a feature map $F \in \mathbb{R}^{n \times n \times t}$ is discretized from the smooth function $e : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$, $|S| = t$, the mesh size is h , $S(\cdot)$ is the stride downsampling operator. If for any $A, B \in S, x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\|\nabla e(x, A)\| \leq G, \quad (17)$$

then the following results are satisfied:

$$|S[\tilde{\pi}_B^E](F)(x, A) - \pi_B^E[S(F)](x, A)| \leq 2\sqrt{2}Gh. \quad (18)$$

Proof. From the above definition we can deduce as follows:

$$[S[\tilde{\pi}_B^E](F)](x, A) = S(\tilde{\pi}_B^E(F))(x, A) = (\tilde{\pi}_B^E(F))_{i, j+1}^A = e(B^{-1}x_{i, j+1}, B^{-1}A), \quad (19)$$

$$\pi_B^E[M(F)](x, A) = M(F)(B^{-1}x, B^{-1}A) = e(x_{i', j'+1}, B^{-1}A), \quad (20)$$

where $x = [x_1, x_2]^T \in \mathbb{R}^2$, $B^{-1}x = [x'_1, x'_2]^T \in \mathbb{R}^2$ denotes the spatial coordinates, and $x_1 \in [x_{ij_1}, x_{i+1, j_1}]$, $x_2 \in [x_{ij_2}, x_{i, j_2+1}]$, $x'_1 \in [x_{i'j'_1}, x_{i'+1, j'_1}]$, $x'_2 \in [x_{i'j'_2}, x_{i', j'_2+1}]$. Similarly, we have

$$\|B^{-1}x_{i, j+1} - x_{i', j'+1}\|_2^2 \leq \|B^{-1}x_{i, j+1} - B^{-1}x\|_2^2 + \|x_{i', j'+1} - B^{-1}x\|_2^2 \leq \sqrt{2}h + \sqrt{2}h = 2\sqrt{2}h. \quad (21)$$

From (17) and (21) we can derive:

$$|S[\tilde{\pi}_B^E](F)(x, A) - \pi_B^E[S(F)](x, A)| = |e(B^{-1}x_{i, j+1}, B^{-1}A) - e(x_{i', j'+1}, B^{-1}A)| \leq 2\sqrt{2}Gh. \quad (22)$$

□

III. PROOF OF EQUIVARIANCE ERROR FOR UPSAMPLING OPERATORS

A. Proof of equivariance error for nearest neighbor upsampling

Theorem 3. Assume that a feature map $F \in \mathbb{R}^{n \times n \times t}$ is discretized from the smooth function $e : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$, $|S| = t$, the mesh size is h , $N(\cdot)$ is the nearest neighbor upsampling operator. If for any $A, B \in S, x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\|\nabla e(x, A)\| \leq G, \quad (23)$$

then the following results are satisfied:

$$|[N[\tilde{\pi}_B^E](F)](x, A) - \pi_B^E[N(F)](x, A)| \leq \sqrt{2}Gh. \quad (24)$$

Proof. From the above definition we can deduce as follows:

$$[N[\tilde{\pi}_B^E](F)](x, A) = [N(\tilde{\pi}_B^E(F))](x, A) = (\tilde{\pi}_B^E(F))_{i\hat{j}}^A = e(B^{-1}x_{i\hat{j}}, B^{-1}A), \quad (25)$$

where $(\hat{i}, \hat{j}) = \arg \min_{ij} \|x_{ij} - x\|_2^2$.

$$\pi_B^E[N(F)](x, A) = [N(F)](B^{-1}x, B^{-1}A) = F_{\hat{i}\hat{j}}^{B^{-1}A} = e(x_{\hat{i}\hat{j}}, B^{-1}A), \quad (26)$$

where $(\tilde{i}, \tilde{j}) = \arg \min_{ij} \|x_{ij} - B^{-1}x\|_2^2$.

From (25) and (26), by using the Lagrange Mean Value Theorem, we can derive

$$|[N[\tilde{\pi}_B^E](F)](x, A) - \pi_B^E[N(F)](x, A)| = |e(B^{-1}x_{\tilde{i}\tilde{j}}, B^{-1}A) - e(x_{\tilde{i}\tilde{j}}, B^{-1}A)| \leq G \|B^{-1}x_{\tilde{i}\tilde{j}} - x_{\tilde{i}\tilde{j}}\|_2^2. \quad (27)$$

By the definition of $x_{\tilde{i}\tilde{j}}$, we have $\|x_{\tilde{i}\tilde{j}} - x\|_2 \leq \frac{\sqrt{2}}{2}h$. Since B is an orthogonal matrix,

$$\|B^{-1}x_{\tilde{i}\tilde{j}} - B^{-1}x\|_2 \leq \frac{\sqrt{2}}{2}h. \quad (28)$$

By the definition of $x_{\tilde{i}\tilde{j}}$,

$$\|x_{\tilde{i}\tilde{j}} - B^{-1}x\|_2 \leq \frac{\sqrt{2}}{2}h. \quad (29)$$

From (28) and (29) we can derive

$$\|B^{-1}x_{\tilde{i}\tilde{j}} - x_{\tilde{i}\tilde{j}}\|_2^2 \leq \|B^{-1}x_{\tilde{i}\tilde{j}} - B^{-1}x\|_2^2 + \|x_{\tilde{i}\tilde{j}} - B^{-1}x\|_2^2 \leq \frac{\sqrt{2}}{2}h + \frac{\sqrt{2}}{2}h = \sqrt{2}h. \quad (30)$$

Combining (27) and (30), we can get

$$|[N[\tilde{\pi}_B^E](F)](x, A) - \pi_B^E[N(F)](x, A)| \leq \sqrt{2}Gh. \quad (31)$$

□

B. Proof of equivariance error for bilinear upsampling

Theorem 4. Assume that a feature map $F \in \mathbb{R}^{n \times n \times t}$ is discretized from the smooth function $e : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$, $|S| = t$, the mesh size is h , $B(\cdot)$ is the bilinear upsampling operator. If for any $A, C \in S$, $x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\|\nabla e(x, A)\| \leq G, \quad (32)$$

then the following results are satisfied:

$$|[B[\tilde{\pi}_C^E](F)](x, A) - \pi_C^E[B(F)](x, A)| \leq 2(\sqrt{2} + 1)Gh. \quad (33)$$

Proof. From the above definition we can deduce as follows:

$$\begin{aligned} [B[\tilde{\pi}_C^E](F)](x, A) &= [B(\tilde{\pi}_C^E(F))](x, A) \\ &= \lambda_1(\tilde{\pi}_C^E(F))_{i,j}^A + \lambda_2(\tilde{\pi}_C^E(F))_{i+1,j}^A + \lambda_3(\tilde{\pi}_C^E(F))_{i,j+1}^A + \lambda_4(\tilde{\pi}_C^E(F))_{i+1,j+1}^A \\ &= \lambda_1 e(C^{-1}x_{ij}, C^{-1}A) + \lambda_2 e(C^{-1}x_{i+1,j}, C^{-1}A) + \lambda_3 e(C^{-1}x_{i,j+1}, C^{-1}A) + \lambda_4 e(C^{-1}x_{i+1,j+1}, C^{-1}A), \end{aligned} \quad (34)$$

$$\begin{aligned} \pi_C^E[B(F)](x, A) &= [B(F)](C^{-1}x, C^{-1}A) \\ &= \lambda'_1 F_{i'j'}^{C^{-1}A} + \lambda'_2 F_{i'+1,j'}^{C^{-1}A} + \lambda'_3 F_{i',j'+1}^{C^{-1}A} + \lambda'_4 F_{i'+1,j'+1}^{C^{-1}A} \\ &= \lambda'_1 e(x_{i'j'}, C^{-1}A) + \lambda'_2 e(x_{i'+1,j'}, C^{-1}A) + \lambda'_3 e(x_{i',j'+1}, C^{-1}A) + \lambda'_4 e(x_{i'+1,j'+1}, C^{-1}A), \end{aligned} \quad (35)$$

where $x = [x_1, x_2]^T \in \mathbb{R}^2$, $C^{-1}x = [x'_1, x'_2]^T \in \mathbb{R}^2$ denotes the spatial coordinates, and $x_1 \in [x_{ij_1}, x_{i+1,j_1}]$, $x_2 \in [x_{ij_2}, x_{i,j+1_2}]$, $x'_1 \in [x_{i'j'_1}, x_{i'+1,j'_1}]$, $x'_2 \in [x_{i'j'_2}, x_{i',j'+1_2}]$. We note that $\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4 = 1$, and they're both positive. We can transform (34) as follows:

$$\begin{aligned} [B[\tilde{\pi}_C^E](F)](x, A) &= \lambda_1 e(C^{-1}x_{ij}, C^{-1}A) + \lambda_2 e(C^{-1}x_{i+1,j}, C^{-1}A) + \lambda_3 e(C^{-1}x_{i,j+1}, C^{-1}A) + \lambda_4 e(C^{-1}x_{i+1,j+1}, C^{-1}A) \\ &\quad - (\lambda_1 + \lambda_2 + \lambda_3 + \lambda_4) e(C^{-1}x_{ij}, C^{-1}A) + e(C^{-1}x_{ij}, C^{-1}A) \\ &= e(C^{-1}x_{ij}, C^{-1}A) + \lambda_2 [e(C^{-1}x_{i+1,j}, C^{-1}A) - e(C^{-1}x_{ij}, C^{-1}A)] \\ &\quad + \lambda_3 [e(C^{-1}x_{i,j+1}, C^{-1}A) - e(C^{-1}x_{ij}, C^{-1}A)] + \lambda_4 [e(C^{-1}x_{i+1,j+1}, C^{-1}A) - e(C^{-1}x_{ij}, C^{-1}A)]. \end{aligned} \quad (36)$$

We can transform (35) in the same way.

$$\begin{aligned}
\pi_C^E[B(F)](x, A) &= \lambda'_1 e(x_{i'j'}, C^{-1}A) + \lambda'_2 e(x_{i'+1,j'}, C^{-1}A) + \lambda'_3 e(x_{i',j'+1}, C^{-1}A) + \lambda'_4 e(x_{i'+1,j'+1}, C^{-1}A) \\
&\quad - (\lambda'_1 + \lambda'_2 + \lambda'_3 + \lambda'_4) e(x_{i'j'}, C^{-1}A) + e(x_{i'j'}, C^{-1}A) \\
&= e(x_{i'j'}, C^{-1}A) + \lambda'_2 [e(x_{i'+1,j'}, C^{-1}A) - e(x_{i'j'}, C^{-1}A)] \\
&\quad + \lambda'_3 [e(x_{i',j'+1}, C^{-1}A) - e(x_{i'j'}, C^{-1}A)] + \lambda'_4 [e(x_{i'+1,j'+1}, C^{-1}A) - e(x_{i'j'}, C^{-1}A)].
\end{aligned} \tag{37}$$

We can find that $\|C^{-1}x_{i+1,j} - C^{-1}x_{ij}\|_2^2 \leq h$, it still holds true for other points. So we have

$$\begin{aligned}
\lambda_2 |e(C^{-1}x_{i+1,j}, C^{-1}A) - e(C^{-1}x_{ij}, C^{-1}A)| + \lambda_3 |e(C^{-1}x_{i,j+1}, C^{-1}A) - e(C^{-1}x_{ij}, C^{-1}A)| \\
+ \lambda_4 |e(C^{-1}x_{i+1,j+1}, C^{-1}A) - e(C^{-1}x_{ij}, C^{-1}A)| \leq Gh,
\end{aligned} \tag{38}$$

$$\begin{aligned}
\lambda'_2 |e(x_{i'+1,j'}, C^{-1}A) - e(x_{i'j'}, C^{-1}A)| + \lambda'_3 |e(x_{i',j'+1}, C^{-1}A) - e(x_{i'j'}, C^{-1}A)| \\
+ \lambda'_4 |e(x_{i'+1,j'+1}, C^{-1}A) - e(x_{i'j'}, C^{-1}A)| \leq Gh,
\end{aligned} \tag{39}$$

$$\|C^{-1}x_{ij} - x_{i'j'}\|_2^2 \leq \|C^{-1}x_{ij} - C^{-1}x\|_2^2 + \|x_{i'j'} - C^{-1}x\|_2^2 \leq \sqrt{2}h + \sqrt{2}h = 2\sqrt{2}h. \tag{40}$$

Then we have

$$\begin{aligned}
|[B[\tilde{\pi}_C^E](F)](x, A) - \pi_C^E[B(F)](x, A)| &\leq |e(C^{-1}x_{ij}, C^{-1}A) - e(x_{i'j'}, C^{-1}A)| + 2Gh \\
&\leq 2(\sqrt{2} + 1)Gh.
\end{aligned} \tag{41}$$

Finally, we have

$$|[B[\tilde{\pi}_C^E](F)](x, A) - \pi_C^E[B(F)](x, A)| \leq 2(\sqrt{2} + 1)Gh. \tag{42}$$

□

IV. PROOF OF EQUIVARIANCE ERROR OF COMPLETE U-NET NETWORK

In this section, we aim to analyze the rotational equivariance error of the complete U-Net network. **Lemma 1.**, using bilinear upsampling as an example, provides a theoretical guarantee that our interpolation method does not increase the gradient values of the feature maps. **Lemma 2.** establishes the range of function values, derivative values, and second-order derivatives for each feature map layer in a rotationally equivariant network. **Lemma 3.** provides the rotational equivariance error for a single-layer, single-channel convolutional layer. **Theorem 5.** presents the rotational equivariance error of the complete U-Net network under discrete angles.

Lemma 1. Assume that a feature map $F \in \mathbb{R}^{n \times n \times t}$ is discretized from the smooth function $e : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$, $|S| = t$, the mesh size is h , $B(\cdot)$ is the bilinear upsampling operator. If for any $A \in S, x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\|\nabla e(x, A)\| \leq G, \tag{43}$$

then we have:

$$[B(F)](x, A) \leq G. \tag{44}$$

Proof. Based on **Rolle's Theorem**: let a function f be continuous on $[a, b]$ and differentiable on (a, b) . If $f(a) = f(b)$, then there exists a point $\xi \in (a, b)$ such that $f'(\xi) = 0$. At the grid points, we have $[B(F)](x, A) = e(x, A)$. Let $F(x) = [B(F)](x, A) - e(x, A)$, x_1, x_2 are grid points and ξ is a point within the range of the grid points. Then $F(x)$ satisfies Rolle's Theorem and can be expressed as:

$$F'(\xi) = \nabla[B(F)](\xi, A) - \nabla e(\xi, A) = 0. \tag{45}$$

Therefore, we can obtain $\nabla[B(F)](x, A) \subset \nabla e(x, A)$, i.e. $[B(F)](x, A) \leq G$. □

Lemma 2. (Fu, 2023 [3]) For an image X with size $h \times w \times c$, and a N-layer rotation equivariant CNN network $\text{CNN}_{eq}(\cdot)$, whose channel number of the i^{th} layer is n_i , 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^{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 $\phi^i : \mathbb{R}^2 \rightarrow \mathbb{R}$, where $i \in \{1, \dots, N\}, c \in \{1, \dots, C\}$, for any $x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\begin{aligned}
|r_c(x)| &\leq F_0, \|\nabla r_c(x)\| \leq G_0, \|\nabla^2 r_c(x)\| \leq H_0, \\
|\phi^i(x)| &\leq F_i, \|\nabla \phi^i(x)\| \leq G_i, \|\nabla^2 \phi^i(x)\| \leq H_i, \\
\forall \|x\| &\geq (p+1)h/2, \phi_i(x) = 0,
\end{aligned} \tag{46}$$

where p is the filter size, h is the mesh size, and ∇ and ∇^2 denote the operators of gradient and Hessian matrix, respectively. Denote

$$e_d^i(x, B) = \begin{cases} \sum_{c, \delta \in \Lambda} \varphi_{cd}^1(B^{-1}\delta) r_c(x - \delta) & \text{if } i = 1, \\ \sum_{c, A, \delta \in \Lambda} \varphi_{Acd}^i(B^{-1}\delta) e_c^{i-1}(x - \delta, BA) & \text{if } i \neq 1, N. \end{cases} \quad (47)$$

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

$$\begin{aligned} |e_d^i(x, B)| &\leq F_0 \mathcal{F}_i, \\ |\nabla e_d^i(x, B)| &\leq \left(\sum_{m=1}^i \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_i, \\ |\nabla^2 e_d^i(x, B)| &\leq \left(\sum_{m=1}^i \frac{H_m F_0}{F_m} + 2 \sum_{l=1}^i \frac{G_l}{F_l} \sum_{m=1}^{l-1} \frac{G_m F_0}{F_m} + 2 \sum_{m=1}^i \frac{G_m G_0}{F_m} + H_0 \right) \mathcal{F}_i, \end{aligned} \quad (48)$$

where $\mathcal{F}_i = \prod_{k=1}^i n_{k-1} p^2 F_k$, $\forall i = 1, 2, \dots, N-1$. \square

Lemma 3. (Xie. 2023 [4]) Assume that an image $X \in \mathbb{R}^{n \times n}$ is discretized from the smooth function $r : \mathbb{R}^2 \rightarrow \mathbb{R}$, a feature map $F \in \mathbb{R}^{n \times n \times t}$ is discretized from the smooth function $e : \mathbb{R}^2 \times S \rightarrow \mathbb{R}$, $|S| = t$, and filters $\tilde{\Psi}$, $\tilde{\Phi}$ and $\tilde{\Upsilon}$ are generated from φ_{in} , φ_{out} and φ_A , $\forall A \in S$, respectively. If for any $A \in S$, $x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\begin{aligned} |r(x)|, |e(x, A)| &\leq F_1, \\ \|\nabla r(x)\|, \|\nabla e(x, A)\| &\leq G_1, \\ \|\nabla^2 r(x)\|, \|\nabla^2 e(x, A)\| &\leq H_1, \\ |\varphi_{in}(x)|, |\varphi_A(x)|, |\varphi_{out}(x)| &\leq F_2, \\ \|\nabla \varphi_{in}(x)\|, \|\nabla \varphi_A(x)\|, \|\nabla \varphi_{out}(x)\| &\leq G_2, \\ \|\nabla^2 \varphi_{in}(x)\|, \|\nabla^2 \varphi_A(x)\|, \|\nabla^2 \varphi_{out}(x)\| &\leq H_2, \\ \forall \|x\| \geq (p+1)h/2, \varphi_{in}(x), \varphi_A(x), \varphi_{out}(x) &= 0, \end{aligned} \quad (49)$$

where p is the filter size, h is the mesh size, and ∇ and ∇^2 denote the operators of gradient and Hessian matrix, respectively, then for any $\theta_k = \{2k\pi/t \mid k = 1, 2, \dots, t\}$, the following results are satisfied:

$$\begin{aligned} \left| \hat{\Psi}[\tilde{\pi}_{\theta_k}](X) - \tilde{\pi}_{\theta_k}[\hat{\Psi}](X) \right| &\leq \frac{C}{2} (p+1)^2 h^2, \\ \left| \hat{\Phi}[\tilde{\pi}_{\theta_k}](F) - \tilde{\pi}_{\theta_k}[\hat{\Phi}](F) \right| &\leq \frac{C}{2} (p+1)^2 h^2 t, \\ \left| \hat{\Upsilon}[\tilde{\pi}_{\theta_k}](F) - \tilde{\pi}_{\theta_k}[\hat{\Upsilon}](F) \right| &\leq \frac{C}{2} (p+1)^2 h^2 t, \end{aligned} \quad (50)$$

where $C = F_1 H_2 + F_2 H_1 + 2G_1 G_2$ and $\|\cdot\|_\infty$ represents the infinity norm. \square

To obtain the rotational equivariance error of the complete U-Net network under discrete angles, as shown in fig. 1, we decompose the network into multiple upsampling and downsampling blocks. Each downsampling block (DB) consists of one E-Conv layer (Equivariant Convolution) and a downsampling operator, while each upsampling block (UB) is composed of an upsampling operator and two E-Conv layers, we provide the equivariant error for each block. These are denoted as **Lemma 4.** and **Lemma 5.** Note that the mesh size of upsampling and downsampling is different, **we define the mesh size of the original picture to be h** , the mesh size after a $\times 2$ downsampling is $2h$, and so on.

Lemma 4.(DB Equivariant Error) We take an equivariant convolution layer $\Phi(\cdot)$ and a downsampling layer $D(\cdot)$ as a Downsampling Block (DB), $\Phi(\cdot)$ and $D(\cdot)$ are both multi-channel operators, we consider the i^{th} DB module DB_i , we used stride downsampling operator for demonstration. 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 $\phi^i : \mathbb{R}^2 \rightarrow \mathbb{R}$, where $l \in \{1, \dots, L\}$, $c \in \{1, \dots, C\}$, for any $x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\begin{aligned} |r_c(x)| &\leq F_0, \|\nabla r_c(x)\| \leq G_0, \|\nabla^2 r_c(x)\| \leq H_0, \\ |\phi^i(x)| &\leq F_i, \|\nabla \phi^i(x)\| \leq G_i, \|\nabla^2 \phi^i(x)\| \leq H_i, \\ \forall \|x\| \geq (p+1)h/2, \varphi_i(x) &= 0, \end{aligned} \quad (51)$$

where p is the filter size, h is the mesh size of the original picture, and ∇ and ∇^2 denote the operators of gradient and Hessian matrix, respectively. In front of this module there are i convolutions ($i-1$ intermediate layers and 1 input layer) and $i-1$ downsampling modules, and we give the rotation equivariant error of the DB_i , the following result is satisfied:

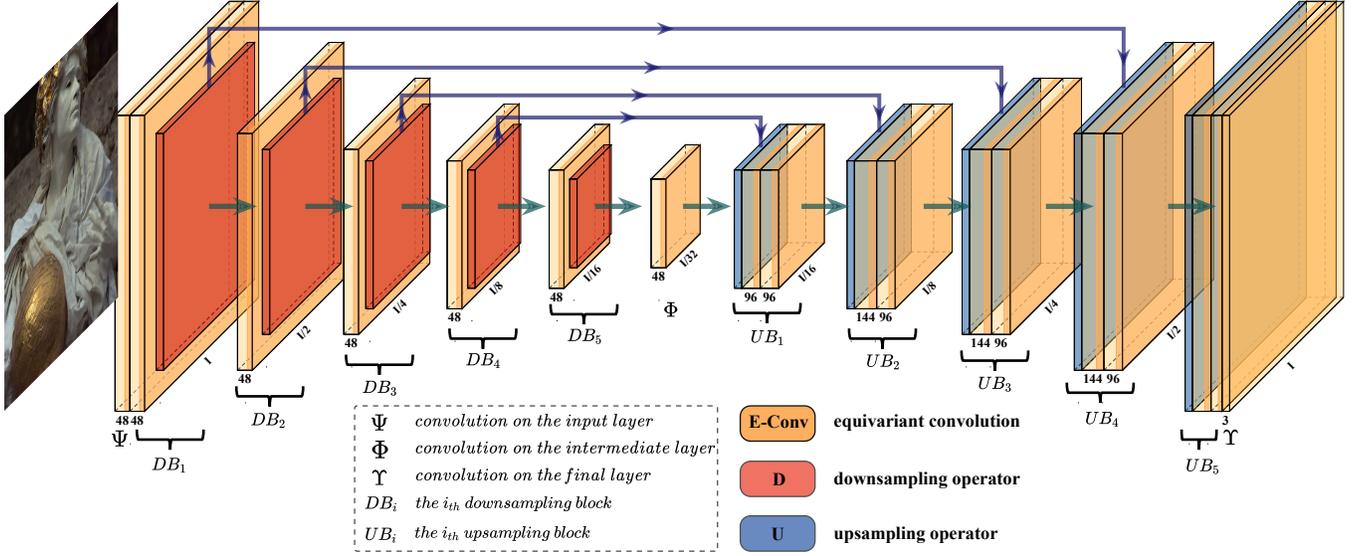


Fig. 1. The network architecture of the equivariant N2N method. The network can be divided into multiple upsampling and downsampling blocks. Each downsampling block (DB) consists of one E-Conv layer and a downsampling operator, while each upsampling block (UB) is composed of an upsampling operator and two E-Conv layers.

$$\left| D\hat{B}_i [\tilde{\pi}_\theta^E] (F)(x) - \pi_\theta^E [D\hat{B}_i(F)] (x) \right| \leq Q_{i1}h + Q_{i2}h^2, \quad (52)$$

where $Q_{i1} = \left(\sum_{m=1}^{i+1} \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_{i+1} 2^{i-\frac{1}{2}}$, $Q_{i2} = \left(\sum_{m=1}^{i+1} \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_{i+1} n_i \frac{C_{i+1}}{2} (p+1)^2 2^{2i-2}$.

Proof.

$$\begin{aligned} & \left| D\hat{B}_i [\tilde{\pi}_\theta^E] (F)(x) - \pi_\theta^E [D\hat{B}_i(F)] (x) \right| = \left| D \left(\hat{\Phi}_{i+1} [\tilde{\pi}_\theta^E] (F) \right) (x) - \pi_\theta^E \left[D \left[\hat{\Phi}_{i+1} \right] (F) \right] (x) \right| \\ & \leq \left| D \left(\hat{\Phi}_{i+1} [\tilde{\pi}_\theta^E] (F) \right) (x) - D \left(\tilde{\pi}_\theta^E \left(\hat{\Phi}_{i+1} \right) (F) \right) (x) \right| + \left| D \left(\tilde{\pi}_\theta^E \left(\hat{\Phi}_{i+1} \right) (F) \right) (x) - \pi_\theta^E \left[D \left[\hat{\Phi}_{i+1} \right] (F) \right] (x) \right|. \end{aligned} \quad (53)$$

By Lemma 2, we know that the derivative of the feature map at the i_{th} layer satisfies:

$$\left| \nabla e_d^i(x, B) \right| \leq \left(\sum_{m=1}^i \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_i, \quad (54)$$

so the derivative of layer $i+1$ is

$$\left| \nabla e_d^{i+1}(x, B) \right| \leq \left(\sum_{m=1}^{i+1} \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_{i+1}. \quad (55)$$

For the second absolute value, since there are $i-1$ downsampling modules before it, the meshsize has become $2^{i-1}h$. We define the meshsize of the original image as h , and we use the derivative of the $i+1_{th}$ layer, then we have:

$$\left| D \left(\tilde{\pi}_\theta^E \left(\hat{\Phi}_{i+1} \right) (F) \right) (x) - \pi_\theta^E \left[D \left[\hat{\Phi}_{i+1} \right] (F) \right] (x) \right| \leq 2\sqrt{2} \left(\sum_{m=1}^{i+1} \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_{i+1} 2^{i-1}h = 2^{i+\frac{1}{2}} \left(\sum_{m=1}^{i+1} \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_{i+1}h.$$

For the first absolute value, according to Lemma 3, we have:

$$\begin{aligned} & \left| D \left(\hat{\Phi}_{i+1} [\tilde{\pi}_\theta^E] (F) \right) (x) - D \left(\tilde{\pi}_\theta^E \left(\hat{\Phi}_{i+1} \right) (F) \right) (x) \right| = \left| D \left(\hat{\Phi}_{i+1} [\tilde{\pi}_\theta^E] (F)(x) - \tilde{\pi}_\theta^E \left[\hat{\Phi}_{i+1} \right] (F)(x) \right) \right| \\ & \leq \left(\sum_{m=1}^{i+1} \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_{i+1} n_i \frac{C_{i+1}}{2} (p+1)^2 (2^{i-1}h)^2. \end{aligned} \quad (56)$$

Then we have:

$$\left| D\hat{B}_i [\tilde{\pi}_\theta^E] (F)(x) - \pi_\theta^E [D\hat{B}_i(F)] (x) \right| \leq Q_{i1}h + Q_{i2}h^2, \quad (57)$$

where $Q_{i1} = \left(\sum_{m=1}^{i+1} \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_{i+1} 2^{i+\frac{1}{2}}$, $Q_{i2} = \left(\sum_{m=1}^{i+1} \frac{G_m F_0}{F_m} + G_0 \right) \mathcal{F}_{i+1} n_i \frac{C_{i+1}}{2} (p+1)^2 2^{2i-2}$. \square

Lemma 5.(UB Equivariant Error) We take an upsampling layer $U(\cdot)$ and two equivariant convolution layers $\Phi_i(\cdot), \Phi_{i+1}(\cdot)$ as a Upsampling Block (UB), all operators are multi-channel, we consider the i^{th} UB module UB_i , we used nearest neighbor upsampling operator for demonstration. 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 $\phi^i : \mathbb{R}^2 \rightarrow \mathbb{R}$, where $l \in \{1, \dots, L\}, c \in \{1, \dots, C\}$, for any $x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\begin{aligned} |r_c(x)| &\leq F_0, \|\nabla r_c(x)\| \leq G_0, \|\nabla^2 r_c(x)\| \leq H_0, \\ |\varphi^i(x)| &\leq F_i, \|\nabla \varphi^i(x)\| \leq G_i, \|\nabla^2 \varphi^i(x)\| \leq H_i, \\ \forall \|x\| &\geq (p+1)h/2, \varphi_i(x) = 0, \end{aligned} \quad (58)$$

where p is the filter size, h is the mesh size of the original picture, and ∇ and ∇^2 denote the operators of gradient and Hessian matrix, respectively. In front of this module there are $m+2+2i-2 = m+2i$ convolutions, m downsampling modules and $i-1$ upsampling modules, and we give the rotation equivariant error of the UB_i , the following result is satisfied:

$$\left| U\hat{B}_i [\tilde{\pi}_\theta^E] (F)(x) - \tilde{\pi}_\theta^E \left[U\hat{B}_i(F) \right] (x) \right| \leq K_{i1}h + K_{i2}h^2, \quad (59)$$

where $K_{i1} = n_{m+2i+1} p^2 F_{m+2i+2} n_{m+2i} p^2 F_{m+2i+1} \sqrt{2} \left(\sum_{j=1}^{m+2i} \frac{G_j F_0}{F_j} + G_0 \right) 2^{m-i+1}$,
 $K_{i2} = n_{m+2i+1} p^2 F_{m+2i+2} n_{m+2i} \frac{C_{m+2i+1}}{2} (p+1)^2 2^{2m-2i} + n_{m+2i+1} \frac{C_{m+2i+2}}{2} (p+1)^2 2^{2m-2i}$.

Proof.

$$\begin{aligned} \left| U\hat{B}_i [\tilde{\pi}_\theta^E] (F)(x) - \tilde{\pi}_\theta^E \left[U\hat{B}_i(F) \right] (x) \right| &= \left| \hat{\Phi}_{m+2i+2} \left(\hat{\Phi}_{m+2i+1} U [\tilde{\pi}_\theta^E] (F) \right) (x) - \tilde{\pi}_\theta^E \left[\hat{\Phi}_{m+2i+2} \hat{\Phi}_{m+2i+1} [U] (F) \right] (x) \right| \\ &\leq \left| \hat{\Phi}_{m+2i+2} \left(\hat{\Phi}_{m+2i+1} U [\tilde{\pi}_\theta^E] (F) \right) (x) - \hat{\Phi}_{m+2i+2} \left(\hat{\Phi}_{m+2i+1} \pi_\theta^E [U] (F) \right) (x) \right| \\ &\quad + \left| \hat{\Phi}_{m+2i+2} \left(\hat{\Phi}_{m+2i+1} \pi_\theta^E [U] (F) \right) (x) - \hat{\Phi}_{m+2i+2} \left(\tilde{\pi}_\theta^E \hat{\Phi}_{m+2i+1} [U] (F) \right) (x) \right| \\ &\quad + \left| \hat{\Phi}_{m+2i+2} \left(\tilde{\pi}_\theta^E \hat{\Phi}_{m+2i+1} [U] (F) \right) (x) - \tilde{\pi}_\theta^E \hat{\Phi}_{m+2i+2} \left(\hat{\Phi}_{m+2i+1} [U] (F) \right) (x) \right|. \end{aligned}$$

We assume there are m DB modules, in the instance m is equal to 5, but for generality, we denote it as m . For the first absolute value, the derivative function is from the $(m+2i)_{\text{th}}$ layer, which has undergone m downsampling and $i-1$ upsampling operations. Therefore, h has been amplified by a factor of $m-i+1$, and we have:

$$\begin{aligned} &\left| \hat{\Phi}_{m+2i+2} \left(\hat{\Phi}_{m+2i+1} U [\tilde{\pi}_\theta^E] (F) \right) (x) - \hat{\Phi}_{m+2i+2} \left(\hat{\Phi}_{m+2i+1} \pi_\theta^E [U] (F) \right) (x) \right| \\ &\leq n_{m+2i+1} p^2 F_{m+2i+2} n_{m+2i} p^2 F_{m+2i+1} \sqrt{2} \left(\sum_{j=1}^{m+2i} \frac{G_j F_0}{F_j} + G_0 \right) 2^{m-i+1} h. \end{aligned} \quad (60)$$

For the second absolute value, applying the mean value theorem for derivatives, this concerns the rotation before and after the $(m+2i+1)_{\text{th}}$ layer, which has undergone m downsampling and i upsampling operations, and we have:

$$\begin{aligned} &\left| \hat{\Phi}_{m+2i+2} \left(\hat{\Phi}_{m+2i+1} \pi_\theta^E [U] (F) \right) (x) - \hat{\Phi}_{m+2i+2} \left(\tilde{\pi}_\theta^E \hat{\Phi}_{m+2i+1} [U] (F) \right) (x) \right| \\ &\leq n_{m+2i+1} p^2 F_{m+2i+2} n_{m+2i} \frac{C_{m+2i+1}}{2} (p+1)^2 (2^{m-i} h)^2. \end{aligned} \quad (61)$$

For the third absolute value, we have:

$$\left| \hat{\Phi}_{m+2i+2} \left(\tilde{\pi}_\theta^E \hat{\Phi}_{m+2i+1} [U] (F) \right) (x) - \tilde{\pi}_\theta^E \hat{\Phi}_{m+2i+2} \left(\hat{\Phi}_{m+2i+1} [U] (F) \right) (x) \right| \leq n_{m+2i+1} \frac{C_{m+2i+2}}{2} (p+1)^2 (2^{m-i} h)^2. \quad (62)$$

Adding the three absolute values together, we obtain the final result:

$$\left| U\hat{B}_i [\tilde{\pi}_\theta^E] (F)(x) - \tilde{\pi}_\theta^E \left[U\hat{B}_i(F) \right] (x) \right| \leq K_{i1}h + K_{i2}h^2, \quad (63)$$

where $K_{i1} = n_{m+2i+1} p^2 F_{m+2i+2} n_{m+2i} p^2 F_{m+2i+1} \sqrt{2} \left(\sum_{j=1}^{m+2i} \frac{G_j F_0}{F_j} + G_0 \right) 2^{m-i+1}$,
 $K_{i2} = n_{m+2i+1} p^2 F_{m+2i+2} n_{m+2i} \frac{C_{m+2i+1}}{2} (p+1)^2 2^{2m-2i} + n_{m+2i+1} \frac{C_{m+2i+2}}{2} (p+1)^2 2^{2m-2i}$. \square

Using the provided Lemma 4 and Lemma 5, we further derive the rotational equivariance error of the complete U-Net network under fixed discrete angles, which is presented as Theorem 5.

Theorem 5. For an image X with size $h \times w \times n_0$, and a rotation equivariant U-Net network $\text{UNet}_{eq}(\cdot)$, whose channel number of the i^{th} layer is n_i , 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^{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 $\phi^i : \mathbb{R}^2 \rightarrow \mathbb{R}$, $\hat{\Psi}$, $\hat{\Phi}$, and $\hat{\Upsilon}$ represent the convolutional layers in the input, middle, and output stages, respectively. We consider m DBs and m UBs, so the network has a total of $3m + 3$ convolutional layers. We define:

$$\text{UNet}_{eq}(\cdot) = \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \cdots \right] \right] \right] (\cdot), \quad (64)$$

the following conditions are satisfied:

$$\begin{aligned} |r_c(x)| &\leq F_0, \|\nabla_x r_c(x)\| \leq G_0, \|\nabla_x^2 r_c(x)\| \leq H_0, \\ |\phi^l(x)| &\leq F_l, \|\nabla_x \phi^l(x)\| \leq G_l, \|\nabla_x^2 \phi^l(x)\| \leq H_l, \\ \forall \|x\| &\geq (p+1)h/2, \phi_l(x) = 0, \end{aligned} \quad (65)$$

where p is the filter size, h is the mesh size, $\theta_k = \frac{2k\pi}{t}$, $k = 1, 2, \dots, t$ and ∇_x and ∇_x^2 denote the operators of gradient and Hessian matrix, respectively. We have

$$|\text{UNet}_{eq}[\tilde{\pi}_{\theta_k}^R](X) - \tilde{\pi}_{\theta_k}^R[\text{UNet}_{eq}(X)]| \leq R_1 h + R_2 h^2, \quad (66)$$

where $R_1 = \sum_{i=1}^m \left(\prod_{k=i+2}^{3m+3} n_{k-1} p^2 F_k \right) Q_{i1} + \sum_{i=1}^m \left(\prod_{k=m+2i+3}^{3m+3} n_{k-1} p^2 F_k \right) K_{i1}$, $R_2 = \left(\prod_{k=2}^{3m+3} n_{k-1} p^2 F_k \right) n_0 \frac{C_1}{2} (p+1)^2 + \sum_{i=1}^m \left(\prod_{k=i+2}^{3m+3} n_{k-1} p^2 F_k \right) Q_{i2} + \left(\prod_{k=m+3}^{3m+3} n_{k-1} p^2 F_k \right) n_{m+1} \frac{C_{m+2}}{2} (p+1)^2 (2^{2m}) + \sum_{i=1}^m \left(\prod_{k=m+2i+3}^{3m+3} n_{k-1} p^2 F_k \right) K_{i2} + n_{3m+2} \frac{C_{3m+3}}{2} (p+1)^2$.

Proof. We assume there are m upsampling modules and m subsampling modules.

$$\begin{aligned} &|\text{UNet}_{eq}[\tilde{\pi}_{\theta_k}^R](X) - \tilde{\pi}_{\theta_k}^R[\text{UNet}_{eq}(X)]| \\ &= \left| \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_1 \left[\hat{\Psi} \left[\tilde{\pi}_{\theta_k}^R \right] \right] \right] \right] \right] \right] (X) - \tilde{\pi}_{\theta_k}^R \left[\hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] (X) \right| \\ &\leq \left| \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_1 \left[\hat{\Psi} \left[\tilde{\pi}_{\theta_k}^R \right] \right] \right] \right] \right] \right] (X) - \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_1 \left[\tilde{\pi}_{\theta_k}^E \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) \right| \\ &+ \left| \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_1 \left[\tilde{\pi}_{\theta_k}^E \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) - \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \left[\tilde{\pi}_{\theta_k}^E \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) \right| \\ &+ \cdots \\ &+ \left| \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_i \left[\tilde{\pi}_{\theta_k}^E \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) - \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \left[\tilde{\pi}_{\theta_k}^E \hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) \right| \\ &+ \cdots \\ &+ \left| \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\tilde{\pi}_{\theta_k}^E \cdots \hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) - \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \tilde{\pi}_{\theta_k}^E \left[\hat{D}B_m \cdots \left[\hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) \right| \\ &+ \left| \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \tilde{\pi}_{\theta_k}^E \left[\hat{D}B_m \left[\cdots \hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) - \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \tilde{\pi}_{\theta_k}^E \hat{\Phi} \left[\hat{D}B_m \cdots \left[\hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) \right| \\ &+ \left| \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_1 \tilde{\pi}_{\theta_k}^E \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) - \hat{\Upsilon} \left[\hat{U}B_m \cdots \tilde{\pi}_{\theta_k}^E \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \left[\hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) \right| \\ &+ \cdots \\ &+ \left| \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_i \tilde{\pi}_{\theta_k}^E \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) - \hat{\Upsilon} \left[\hat{U}B_m \cdots \tilde{\pi}_{\theta_k}^E \hat{U}B_i \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) \right| \\ &+ \left| \hat{\Upsilon} \left[\hat{U}B_m \tilde{\pi}_{\theta_k}^E \cdots \hat{U}B_i \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) - \hat{\Upsilon} \left[\tilde{\pi}_{\theta_k}^E \hat{U}B_m \cdots \hat{U}B_i \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) \right| \\ &+ \left| \hat{\Upsilon} \tilde{\pi}_{\theta_k}^E \left[\hat{U}B_m \cdots \hat{U}B_i \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) - \tilde{\pi}_{\theta_k}^R \hat{\Upsilon} \left[\hat{U}B_m \cdots \hat{U}B_i \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] \right] (X) \right|. \end{aligned}$$

There are a total of $3m + 3$ convolutions. Following the input convolutional layer, there are $m + 1 + 2m + 1 = 3m + 2$ convolutional layers. We do not need to consider the impact of upsampling and downsampling operators on the feature maps, as demonstrated in Lemma 1. The upsampling and downsampling operators perform interpolation between grid points without increasing the gradient of the feature maps. However, we must consider the mesh size change of h . We can then derive the error of the first layer.

For the rotation equivariant error of the input convolutional layer $\hat{\Psi}(\cdot)$.

$$\begin{aligned} & \left| \hat{Y} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_1 \left[\hat{\Psi} \left[\tilde{\pi}_{\theta_k}^R \right] \right] \right] \right] \right] \right] (X) - \hat{Y} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_1 \left[\tilde{\pi}_{\theta_k}^E \left[\hat{\Psi} \right] \right] \right] \right] \right] (X) \right| \\ & \leq \left(\prod_{k=2}^{3m+3} n_{k-1} p^2 F_k \right) n_0 \frac{C_1}{2} (p+1)^2 h^2, \end{aligned}$$

where $C_1 = H_1 F_0 + F_1 H_0 + 2G_1 G_0$.

For the rotation equivariant error of DB_i , there are i convolutional layers and $i-1$ downsampling layers. There are $m-i+1+2m+1=3m-i+2$ convolutional layers, from the $(i+2)_{th}$ convolutional layer to the $(3m+3)_{th}$ convolutional layer.

$$\begin{aligned} & \left| \hat{Y} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \hat{D}B_i \left[\tilde{\pi}_{\theta_k}^E \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] (X) - \hat{Y} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \cdots \left[\tilde{\pi}_{\theta_k}^E \hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] (X) \right| \\ & \leq n_{3m+2} p^2 F_{3m+3} \cdots n_{i+1} p^2 F_{i+2} (Q_{i1} h + Q_{i2} h^2) = \left(\prod_{k=i+2}^{3m+3} n_{k-1} p^2 F_k \right) (Q_{i1} h + Q_{i2} h^2), \end{aligned}$$

and we also need to sum i from 1 to m . i is the index of the DB .

For the rotation equivariant error of the middle convolutional layer $\hat{\Phi}(\cdot)$,

$$\begin{aligned} & \left| \hat{Y} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\hat{\Phi} \tilde{\pi}_{\theta_k}^E \left[\hat{D}B_m \left[\cdots \hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] (X) - \hat{Y} \left[\hat{U}B_m \cdots \hat{U}B_1 \left[\tilde{\pi}_{\theta_k}^E \hat{\Phi} \left[\hat{D}B_m \cdots \left[\hat{D}B_i \cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] (X) \right| \\ & \leq n_{3m+2} p^2 F_{3m+3} \cdots n_{m+2} p^2 F_{m+3} n_{m+1} \frac{C_{m+2}}{2} (p+1)^2 (2^m h)^2 = \left(\prod_{k=m+3}^{3m+3} n_{k-1} p^2 F_k \right) n_{m+1} \frac{C_{m+2}}{2} (p+1)^2 (2^m h)^2. \end{aligned}$$

For the rotation equivariant error of UB_i , there are $2m-2i+1$ convolutions following it, ranging from the $m+2+2i+1 = m+2i+3$ layer to the $3m+3$ layer.

$$\begin{aligned} & \left| \hat{Y} \left[\hat{U}B_m \cdots \hat{U}B_i \tilde{\pi}_{\theta_k}^E \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] (X) - \hat{Y} \left[\hat{U}B_m \cdots \tilde{\pi}_{\theta_k}^E \hat{U}B_i \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] (X) \right| \\ & \leq n_{3m+2} p^2 F_{3m+3} \cdots n_{m+2i+2} p^2 F_{m+2i+3} (K_{i1} h + K_{i2} h^2) = \left(\prod_{k=m+2i+3}^{3m+3} n_{k-1} p^2 F_k \right) (K_{i1} h + K_{i2} h^2), \end{aligned}$$

and we also need to sum i from 1 to m . i is the index of UB .

For the rotation equivariant error of the output convolutional layer $\hat{Y}(\cdot)$,

$$\begin{aligned} & \left| \hat{Y} \tilde{\pi}_{\theta_k}^E \left[\hat{U}B_m \cdots \hat{U}B_i \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] \right] (X) - \tilde{\pi}_{\theta_k}^R \hat{Y} \left[\hat{U}B_m \cdots \hat{U}B_i \cdots \hat{U}B_1 \left[\hat{\Phi} \left[\hat{D}B_m \left[\cdots \hat{D}B_1 \left[\hat{\Psi} \right] \right] \right] \right] (X) \right| \\ & \leq n_{3m+2} \frac{C_{3m+3}}{2} (p+1)^2 h^2. \end{aligned}$$

In the above derivation, by substituting Lemma 2, we can obtain the expansion of C_i , where

$$C_i = \mathcal{F}_{i-1} F_i \left(\frac{H_i F_0}{F_i} + \sum_{m=1}^{i-1} \frac{H_m F_0}{F_m} + 2 \sum_{l=1}^{i-1} \frac{G_l}{F_l} \sum_{m=1}^{l-1} \frac{G_m F_0}{F_m} + 2 \sum_{m=1}^{i-1} \frac{G_m G_0}{F_m} + H_0 + \sum_{m=1}^{i-1} 2 \frac{G_i G_m F_0}{F_i F_m} + 2 \frac{G_i G_0}{F_i} \right). \quad (67)$$

To make it appear more concise and clear, we use C_i as the expression without expanding it.

Finally we can deduce:

$$|\text{UNet}_{eq} [\tilde{\pi}_{\theta_k}^R] (X) - \tilde{\pi}_{\theta_k}^R [\text{UNet}_{eq}] (X)| \leq R_1 h + R_2 h^2, \quad (68)$$

where $R_1 = \sum_{i=1}^m \left(\prod_{k=i+2}^{3m+3} n_{k-1} p^2 F_k \right) Q_{i1} + \sum_{i=1}^m \left(\prod_{k=m+2i+3}^{3m+3} n_{k-1} p^2 F_k \right) K_{i1}$, $R_2 = \left(\prod_{k=2}^{3m+3} n_{k-1} p^2 F_k \right) n_0 \frac{C_1}{2} (p+1)^2 + \sum_{i=1}^m \left(\prod_{k=i+2}^{3m+3} n_{k-1} p^2 F_k \right) Q_{i2} + \left(\prod_{k=m+3}^{3m+3} n_{k-1} p^2 F_k \right) n_{m+1} \frac{C_{m+2}}{2} (p+1)^2 (2^{2m}) + \sum_{i=1}^m \left(\prod_{k=m+2i+3}^{3m+3} n_{k-1} p^2 F_k \right) K_{i2} + n_{3m+2} \frac{C_{3m+3}}{2} (p+1)^2$. \square

Next, we will derive the rotation equivariant error of the entire U-Net network under any angle. We decompose the equation into the following form (69). The error of the second term has already been provided in Theorem 5. For the errors of the first and third terms, since upsampling and downsampling operations only affect the size of h , we can similarly use the lemma from [3]. For clearer presentation, we write them as Lemma 6 and Lemma 7:

$$\leq \underbrace{|\text{UNet}_{eq} [\tilde{\pi}_{\theta}^R] (X) - \tilde{\pi}_{\theta}^R [\text{UNet}_{eq}] (X)|}_{(1)} + \underbrace{|\text{UNet}_{eq} [\tilde{\pi}_{\theta_k}^R] (X) - \text{UNet}_{eq} [\tilde{\pi}_{\theta_k}^R] (X)|}_{(2)} + \underbrace{|\tilde{\pi}_{\theta_k}^R [\text{UNet}_{eq}] (X) - \tilde{\pi}_{\theta}^R [\text{UNet}_{eq}] (X)|}_{(3)}, \quad (69)$$

Lemma 6. (Fu, 2023 [3]) For an image X with size $H \times W \times n_0$, and a N -layer rotation equivariant network $\text{CNN}_{eq}(\cdot)$, whose channel number of the i^{th} layer is n_i , 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^{th} channel of X 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 $\varphi^i : \mathbb{R}^2 \rightarrow \mathbb{R}$, where $i \in \{1, \dots, N\}$, $c \in \{1, \dots, n_0\}$, for any $x \in \mathbb{R}^2$, the following conditions are satisfied:

$$\begin{aligned} |r_c(x)| &\leq F_0, \|\nabla r_c(x)\| \leq G_0, \|\nabla^2 r_c(x)\| \leq H_0, \\ |\varphi^i(x)| &\leq F_i, \|\nabla \varphi^i(x)\| \leq G_i, \|\nabla^2 \varphi^i(x)\| \leq H_i, \\ \forall \|x\| \geq (p+1)h/2, \varphi_i(x) &= 0, \end{aligned}$$

where p is the filter size, h is the mesh size, and ∇ and ∇^2 denote the operators of gradient and Hessian matrix, respectively. For an arbitrary $\theta \in [0, 2\pi]$, A_{θ} denotes the rotation matrix. If $F(\theta) = \text{CNN}_{eq} [\tilde{\pi}_{\theta}^R] (X) = \hat{\Upsilon} \left[\hat{\Phi}_{N-1} \cdots \hat{\Phi}_{i+1} \left[\hat{\Phi}_i \cdots \hat{\Phi}_2 \left[\hat{\Psi} [\tilde{\pi}_{\theta}^R] \right] \cdots \right] \right] (X)$, then the following result is satisfied:

$$|F'(\theta)| \leq \mathcal{F}(\max\{H, W\} + N(p+1))hG_0,$$

where $\mathcal{F} = \prod_{k=1}^N n_{k-1} p^2 F_k$.

Lemma 7. (Fu, 2023 [3]) Under the same conditions with Lemma 6.

If $F(\theta) = \tilde{\pi}_{\theta}^R [\text{CNN}_{eq}] (X) = \tilde{\pi}_{\theta}^R \left[\hat{\Upsilon} \left[\hat{\Phi}_{N-1} \cdots \hat{\Phi}_{i+1} \left[\hat{\Phi}_i \cdots \hat{\Phi}_2 [\hat{\Psi}] \cdots \right] \right] \right] (X)$, and then the following result is satisfied:

$$|F'(\theta)| \leq \mathcal{F} \max\{H, W\} h G_0,$$

where $\mathcal{F} = \prod_{k=1}^N n_{k-1} p^2 F_k$.

Based on the above two lemmas, we can easily derive the values of the first and third terms. The rotational equivariant error of the complete U-Net network under arbitrary angles is presented in the form of the following corollary:

Corollary 1. Under the same condition as Theorem 5, for an arbitrary $\theta \in [0, 2\pi]$, let π_{θ} denote the rotation transformation, then $\forall \theta$ we have

$$|\text{UNet}_{eq} [\tilde{\pi}_{\theta}^R] (X) - \tilde{\pi}_{\theta}^R [\text{UNet}_{eq}] (X)| \leq R_1 h + R_2 h^2 + R_3 t^{-1} h, \quad (70)$$

where $R_1 = \sum_{i=1}^m \left(\prod_{k=i+2}^{3m+3} n_{k-1} p^2 F_k \right) Q_{i1} + \sum_{i=1}^m \left(\prod_{k=m+2i+3}^{3m+3} n_{k-1} p^2 F_k \right) K_{i1}$, $R_2 = \left(\prod_{k=2}^{3m+3} n_{k-1} p^2 F_k \right) n_0 \frac{C_1}{2} (p+1)^2 + \sum_{i=1}^m \left(\prod_{k=i+2}^{3m+3} n_{k-1} p^2 F_k \right) Q_{i2} + \left(\prod_{k=m+3}^{3m+3} n_{k-1} p^2 F_k \right) n_{m+1} \frac{C_{m+2}}{2} (p+1)^2 (2^{2m}) + \sum_{i=1}^m \left(\prod_{k=m+2i+3}^{3m+3} n_{k-1} p^2 F_k \right) K_{i2} + n_{3m+2} \frac{C_{3m+3}}{2} (p+1)^2$, $R_3 = 2\pi \mathcal{F}(\max\{H, W\} + N(p+1))G_0 + 2\pi \mathcal{F} \max\{H, W\} G_0$.

V. MORE DETAILS ABOUT OUR EXPERIMENTS

A. Datasets and Experimental Indicators

We try our best to conduct experiments in accordance with the original setup and validate our method in three classic approaches. We closely adhered to the official experimental protocol outlined in the reference paper. For the Noise2Noise [5] experiments, we conducted experiments on both RED30 [6] and U-Net networks. U-Net architecture can significantly speed up training, although at the expense of some performance degradation. To improve efficiency, we demonstrated the effectiveness of our method using the U-Net architecture. Our training dataset is the IMAGENET [7] validation set, which comprises 50,000 images. Testing was performed on well-established public datasets: Kodak24 [8], BSDS300 [9], Set14 [10]. In the Noise2Void [11] experiments, we conducted experiments on both grayscale and color images. The R2R [12] experiments were also conducted in accordance with the original settings. Throughout these experiments, we employed PSNR and SSIM as the primary metrics for evaluating image denoising performance.

B. More Experimental Results

In the N2N method, we also used the RED30 network. RED30 has more layers than the U-Net network used in the main text, and its residual structure is more suitable for image denoising tasks. The results are shown in table I. For the N2V method, we followed the original paper's setup and conducted experiments on grayscale images, with the results shown in table II.

| Gaussian25 | | | | | | |
|------------|-----------------|-----------------|-----------------|-----------------|---------------------|-----------------|
| Dataset | N2N [5] | | N2N-EQ | | N2N-EQ ⁺ | |
| | PSNR \uparrow | SSIM \uparrow | PSNR \uparrow | SSIM \uparrow | PSNR \uparrow | SSIM \uparrow |
| Kodak [8] | 31.33 | 0.869 | 31.95 | 0.884 | 31.95 | 0.885 |
| BSD300 [9] | 30.06 | 0.865 | 30.49 | 0.876 | 30.60 | 0.879 |
| Set14 [10] | 29.91 | 0.848 | 30.42 | 0.860 | 30.41 | 0.860 |

| Gaussian50 | | | | | | |
|------------|-----------------|-----------------|-----------------|-----------------|---------------------|-----------------|
| Dataset | N2N [5] | | N2N-EQ | | N2N-EQ ⁺ | |
| | PSNR \uparrow | SSIM \uparrow | PSNR \uparrow | SSIM \uparrow | PSNR \uparrow | SSIM \uparrow |
| Kodak [8] | 27.98 | 0.762 | 28.97 | 0.802 | 28.99 | 0.803 |
| BSD300 [9] | 26.77 | 0.749 | 27.47 | 0.779 | 27.56 | 0.783 |
| Set14 [10] | 26.91 | 0.755 | 27.72 | 0.786 | 27.72 | 0.786 |

TABLE I

N2N: THE THREE NETWORKS WITH RED30 ARCHITECTURE WERE TESTED UNDER CONDITIONS OF GAUSSIAN NOISE AT LEVELS 25 AND 50. THE BEST RESULTS HAVE BEEN BOLDED TO HIGHLIGHT THEM.

| Dataset | σ | N2V [11] | N2V-EQ | N2V-EQ ⁺ |
|---------|----------|-------------|-------------|---------------------|
| BSD68 | 25 | 25.62/0.743 | 26.07/0.726 | 27.47/0.773 |

TABLE II

N2V: EXPERIMENTS ON GRAYSCALE IMAGES. THE BEST RESULTS HAVE BEEN BOLDED TO HIGHLIGHT THEM.

We further design experiments to calculate the equivariant error¹ for the N2V framework, as shown in table III. It can be observed that N2V-EQ with strictly rotation-equivariance achieves the smallest rotational equivariant error. N2V-EQ⁺, relaxes the equivariance constraints, which also achieves smaller equivariant errors than the original network. This verifies that the improvements are achieved by reducing equivariant errors.

| Method | N2V | N2V-EQ | N2V-EQ ⁺ |
|-------------------|--------|--------|---------------------|
| Equivariant Error | 0.2325 | 0.0682 | 0.0759 |

TABLE III

QUANTITATIVE COMPARISON OF ROTATIONAL EQUIVARIANT ERRORS.

We have conducted experiments with the SOTA method in the field of self-supervised fluorescence microscopy denoising [13]. The proposed method achieves a comparable performance with fewer parameters as shown in table IV.

| Method | Confocal | Two-Photon | Wide-Field | #Param |
|---------------|---------------------|---------------------|--------------------|--------|
| | PSNR/SSIM | PSNR/SSIM | PSNR/SSIM | |
| BM3D | 31.28/0.820 | 32.20/0.879 | 31.26/0.760 | - |
| DIP | 33.94/0.902 | 32.35/0.853 | 28.57/0.603 | 2.2M |
| ZSN2N | 34.27/ 0.910 | 32.76/0.869 | 25.84/0.398 | 22.2k |
| FM2S | 34.99/0.909 | 33.46/ 0.883 | 30.14/0.707 | 3.2k |
| FM2S-EQ(ours) | 35.04/0.905 | 33.49/0.881 | 29.93/0.705 | 1.6k |

TABLE IV

DENOISING PERFORMANCE OF FM2S AND DIFFERENT METHODS ON THE NOISY IMAGES FROM THE FMD OFFICIAL TEST SET.

We have supplemented table V with experiments conducted on the real-world datasets.

¹ $\|L_R\Phi(f) - \Phi L_R(f)\|_2^2 / \|L_R\Phi(f)\|_2^2$, where $\Phi(\cdot)$ represents the network and $L_R(\cdot)$ denotes the rotation transformation.

| Method | CBM3D | MCWNNM | N2V | N2S | R2R | R2R-EQ | R2R-EQ ⁺ |
|-----------------|-------|--------|-------|-------|-------|--------------|---------------------|
| SIDD Validation | 25.65 | 33.40 | 29.35 | 30.72 | 34.31 | 34.73 | 34.65 |
| SIDD Benchmark | 25.65 | 33.37 | 27.68 | 29.56 | 35.18 | 35.27 | 34.98 |
| CC | - | - | - | - | 34.31 | 35.28 | 34.34 |
| PolyU | - | - | - | - | 36.44 | 36.69 | 36.35 |
| #Param | - | - | - | 0.22M | 0.67M | 0.17M | 0.84M |

TABLE V
PSNR COMPARISON OF DIFFERENT METHODS.

C. Remark

The difference compared to data augmentation and the necessity and significance of the proposed architecture:

Data augmentation (DA) is a classic way to improve network equivariance. However, the supervision of DA is imposed only on the final output, with no direct impact on the intermediate layers of the network. Therefore, recent research has taken more interest in incorporating equivariance into network architectures. Our approach aims to construct networks with adaptive equivariance, and our sub-networks ensure that all network layers are equivariant. This is significantly different from DA.

Moreover, DA does not conflict with our approach, i.e., embedding rotation equivariance into the model. Actually, further improvements can be achieved with the proposed method, even with DA.

Effectiveness is greater in N2V than in N2N and R2R:

As explained in the manuscript, the success of self-supervised deep learning relies on two factors: 1) prior information from the training data and 2) prior information inherently embedded in the network architecture. In our experiments, N2N and R2R methods require paired noisy images during training, whereas the N2V method can only utilize unpaired noisy images. This means that less prior information can be obtained from the training data in the N2V framework. Therefore, the performance of N2V relies more heavily on the prior information embedded in the network, while the proposed approach exactly focuses on design networks embedded with more rotation symmetry prior. This is why the effectiveness of the proposed method is greater in N2V.

D. More Visualization Results

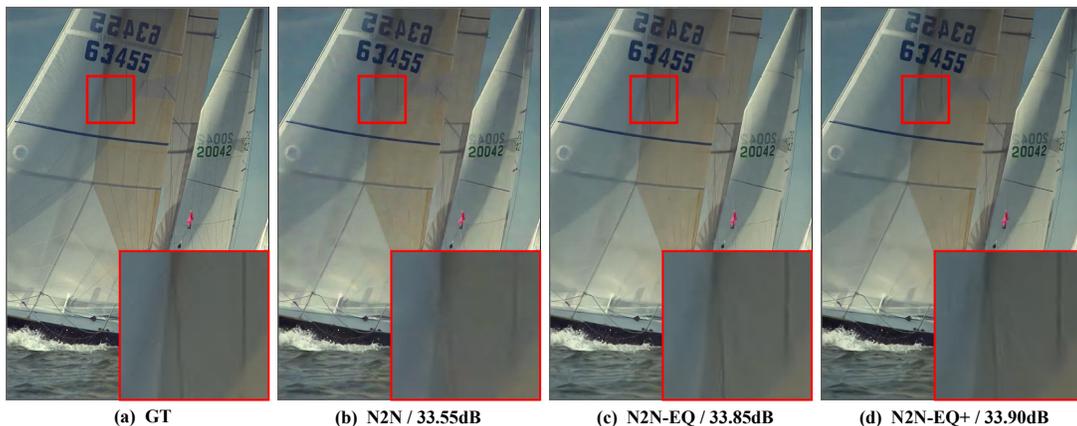


Fig. 2. N2N: Image denoising results of one image from kodak with $\sigma = 25$.

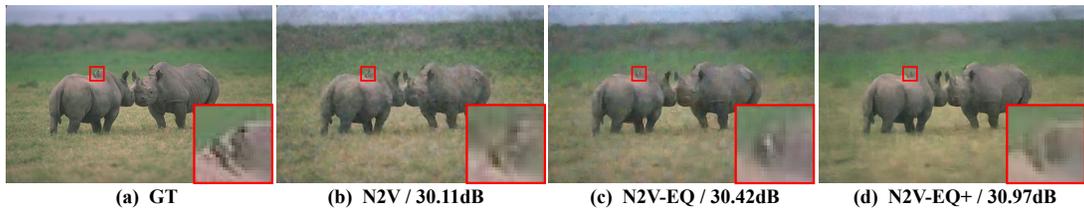


Fig. 3. N2V: Image denoising results of one image from BSD500 with $\sigma = 50$.



Fig. 4. The output result of the MaskNetwork branch in AdaReNet.

REFERENCES

- [1] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.
- [2] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [3] J. Fu, Q. Xie, D. Meng, and Z. Xu, "Rotation equivariant proximal operator for deep unfolding methods in image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.
- [4] Q. Xie, Q. Zhao, Z. Xu, and D. Meng, "Fourier series expansion based filter parametrization for equivariant convolutions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4537–4551, 2022.
- [5] J. Lehtinen, "Noise2noise: Learning image restoration without clean data," *arXiv preprint arXiv:1803.04189*, 2018.
- [6] X.-J. Mao, "Image restoration using convolutional auto-encoders with symmetric skip connections," *arXiv preprint arXiv:1606.08921*, 2016.
- [7] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE conference on computer vision and pattern recognition*. Ieee, 2009, pp. 248–255.
- [8] R. Franzen, "Kodak lossless true color image suite," source: <http://r0k.us/graphics/kodak>, vol. 4, no. 2, p. 9, 1999.
- [9] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, vol. 2. IEEE, 2001, pp. 416–423.
- [10] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7*. Springer, 2012, pp. 711–730.

- [11] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void-learning denoising from single noisy images," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 2129–2137.
- [12] T. Pang, H. Zheng, Y. Quan, and H. Ji, "Recorrupted-to-recorrupted: unsupervised deep learning for image denoising," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021, pp. 2043–2052.
- [13] J. Liu, Q. Teng, and J. Jiang, "Fm2s: Self-supervised fluorescence microscopy denoising with single noisy image," *arXiv preprint arXiv:2412.10031*, 2024.