Appendix A

In this appendix, we will provide a detailed introduction to the impact of interpolation on rotation equivariance.

Supposing that the image I' is the result of I after rotation process R, i.e. $I' = R[I]$, we can establish their relationship by an interpolation process as follow

$$I'\left(\left[\begin{array}{c} x_i \\ y_i \end{array}\right]\right) = R[I]\left(x_i, y_i\right) = I\left(\left[\begin{array}{cc} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{array}\right]\left[\begin{array}{c} x_i \\ y_i \end{array}\right]\right) = \sum_j a_j^i I(x_j, y_j), \tag{A-1}$$

where (x_i, y_i) represents a specific pixel point, where i represents pixel index. α is rotation degree, and a_j^i is the interpolation coefficient, where index j has iterated over all the pixels in the image. We regard the network for semantic segmentation as a combination of multivariate mappings, named f. If we want network f to own rotation equivariance, the following formular should be satisfied:

$$f\left(I'\right) = R[f\left(I\right)]. \tag{A-2}$$

We apply the concept of Currying by treating f as a univariate function simply and examine the change that occurs at the i-th pixel, that is

$$f_i\left(\sum_j a_j^i I(x_j, y_j)\right) = \sum_j a_j^i f_j\left(I(x_j, y_j)\right). \tag{A-3}$$

If we want to satisfy the above equation in any case, f needs to be a linear mapping and $f_i = f_j$ for any j. However, f is always highly nonlinear in almost all networks. Clearly, the trivial solution to this equation should be the case where a_j^i is Kronecker function when f is nonlinear. This corresponds to integer multiples of a 90-degree image rotation or the use of nearest-neighbor interpolation. However, it is generally unlikely that the nearest-neighbor interpolation will select the same neighboring pixels for both input and output, which may introduce greater errors.

As shown in the Fig. A-1, we have the original image, followed by the results of rotating it by 45° using nearest-neighbor interpolation and bilinear interpolation, respectively. The network we employed solely consists of convolutional layers and ReLU activations. Furthermore, the weights within each convolutional kernel are identical, theoretically endowing the network with complete equivariance. However, as evident from the results, neither the outcome from a single layer nor that from five layers demonstrates equivariance under a 45° rotation.

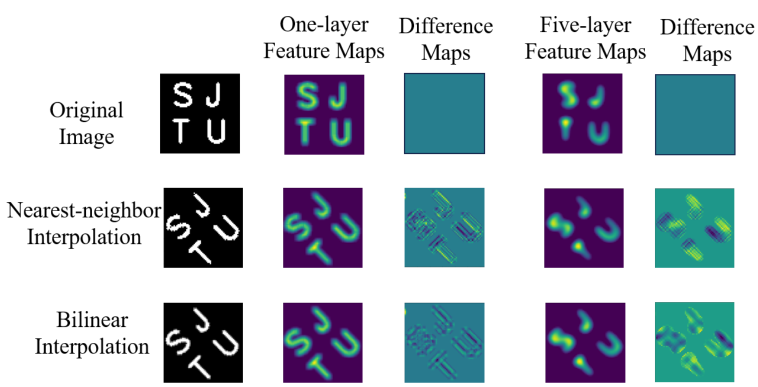

Fig. A-1. Output Feature Maps and Difference Maps after passing through One-Layer and Five-Layer Equivariant Networks under Nearest-neighbor Interpolation and Bilinear Interpolation. Here, the Difference Map specifically refers to the difference between the output feature maps at 45° and 0° (i.e., no rotation).

Appendix B

In this appendix, we will prove the rotation equivariance of proposed convolution-group framework. Regarding the convolutionally distributive law, when applying our specific convolution mode, it can be expressed as:

$$\left(r_{\sigma_t}\varphi\right) \otimes_{\sigma_t} \left(r_{\sigma_t} f\right) = r_{\sigma_t}(\varphi \otimes_{\sigma_0} f) \tag{B-1}$$

where $\sigma_t \in \sigma$, σ_0 is the first one of the cyclic groups. For the convenience of expression, we here give the framework of convolution-group in Fig. B-1.

From Fig. B-1, the feature maps after the first layer are

$$\begin{aligned} \tilde{f}_{\sigma_i}^{2,0} &= \left(r_{\sigma_i}\varphi^1\right) \otimes_{\sigma_i} \left(\tilde{f}^1\right) = \left(r_{\sigma_i}\varphi^1\right) \otimes_{\sigma_i} \left(r_{\sigma_t} f^1\right) \\ &= r_{\sigma_t}\left(r_{\sigma_i} r_{\sigma_t^{-1}}\varphi^1\right) \otimes_{\sigma_i\sigma_t^{-1}} \left(f^1\right) = r_{\sigma_t} f_{\sigma_i\sigma_t^{-1}}^{2,0}. \end{aligned} \tag{B-2}$$

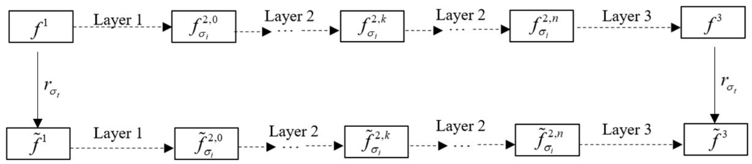

Fig. B-1. The framework of convolution-group. Here, we have r_{σ_t}, and $\sigma_t \in \sigma$. f^1 is the input feature map, and $\tilde{f}^1$ is the feature map after performing some rotation operation on the input feature map. $\tilde{f}^{2,k}_{\sigma_i}$ represent the feature maps obtained by applying the second layer to f^1 at the k-th (k=0,1,2,3...n) time, and $\tilde{f}^{2,k}_{\sigma_i}$ represent the feature maps obtained by applying the second layer to $\tilde{f}^1$ at the k-th time. f^3 represents the output feature map obtained after f^1 passes through the network, and $\tilde{f}^3$ represents the output feature map obtained after $\tilde{f}^1$ passes through the network.

Further, the feature maps after the second layer are

$$\begin{aligned}
\tilde{f}^{2,k+1}_{\sigma_i} &= \sum_{\sigma_i\in\sigma} (r_{\sigma_j}\varphi^{2,k+1}_{\sigma_j^{-1}\sigma_i}) \otimes_{\sigma_i} (\tilde{f}^{2,k}_{\sigma_i}) \\
&= \sum_{\sigma_i\in\sigma} (r_{\sigma_j}\varphi^{2,k+1}_{\sigma_j^{-1}\sigma_i}) \otimes_{\sigma_i} (r_{\sigma_t} f^{2,k}_{\sigma_i\sigma_t^{-1}}) \\
&\xrightarrow{\sigma_i=\sigma_i\sigma_t} \sum_{\sigma_i\in\sigma} (r_{\sigma_j}\varphi^{2,k+1}_{\sigma_j^{-1}\sigma_i\sigma_t}) \otimes_{\sigma_i\sigma_t} (r_{\sigma_t} f^{2,k}_{\sigma_i}) \\
&= \sum_{\sigma_i\in\sigma} r_{\sigma_t}((r_{\sigma_j} r_{\sigma_t^{-1}}\varphi^{2,k+1}_{\sigma_j^{-1}\sigma_i\sigma_t}) \otimes_{\sigma_i} (f^{2,k}_{\sigma_i})) \\
&= r_{\sigma_t}\sum_{\sigma_i\in\sigma} (r_{\sigma_j} r_{\sigma_t^{-1}}\varphi^{2,k+1}_{\sigma_j^{-1}\sigma_i\sigma_t}) \otimes_{\sigma_i} (f^{2,k}_{\sigma_i}) \\
&= r_{\sigma_t} f^{2,k+1}_{\sigma_j\sigma_t^{-1}},
\end{aligned} \tag{B-3}$$

and the feature map after the third layer is

$$\begin{aligned}
\tilde{f}^3 &= \sum_{\sigma_i\in\sigma} (r_{\sigma_i}\varphi^3) \otimes_{\sigma_i} (\tilde{f}^{2,n}_{\sigma_i}) \\
&= \sum_{\sigma_i\in\sigma} (r_{\sigma_i}\varphi^3) \otimes_{\sigma_i} (r_{\sigma_t} f^{2,n}_{\sigma_i\sigma_t^{-1}}) \\
&= r_{\sigma_t}\sum_{\sigma_i\in\sigma} (r_{\sigma_i\sigma_t^{-1}}\varphi^3) \otimes_{\sigma_i\sigma_t^{-1}} (f^{2,n}_{\sigma_i\sigma_t^{-1}}) \\
&\xrightarrow{\sigma_i=\sigma_i\sigma_t} r_{\sigma_t}\sum_{\sigma_i\in\sigma} (r_{\sigma_i}\varphi^3) \otimes_{\sigma_i} (f^{2,n}_{\sigma_i}) \\
&= r_{\sigma_t} f^3.
\end{aligned} \tag{B-4}$$

Clearly, it is seen that after the rotation transformation of input feature map, the output feature map really does the same ordered rotation transformation. Therefore, we can claim that the proposed convolution-group framework is rotation equivariant mathematically.

APPENDIX C

In this appendix, we will deduce Eq. (21a) according to Eq.(20) (i.e., $\Phi(R_{\sigma_t,out}(m), R_{\sigma_t,in}(n)) = \Phi(R_{\sigma_t,\varphi}(m,n))$). We first define

$$\begin{aligned}
&A_d = \tfrac{h_d+w_d}{2}, \Delta A_d = \tfrac{h_d-w_d}{2}, \\
&A_s = \tfrac{h_s+w_s}{2}, \Delta A_s = \tfrac{h_s-w_s}{2}, \\
&A_{in} = \tfrac{h'_{in}+w'_{in}}{2}, \Delta A_{in} = \tfrac{h'_{in}-w'_{in}}{2}, \\
&A_{out} = \tfrac{h_{out}+w_{out}}{2}, \Delta A_{out} = \tfrac{h_{out}-w_{out}}{2}, \\
&A_\varphi = \tfrac{h_\varphi+w_\varphi}{2}, \Delta A_\varphi = \tfrac{h_\varphi-w_\varphi}{2}.
\end{aligned} \tag{C-1}$$

Then, it is easy to get

$$\begin{aligned}
&h_{\sigma_t,d} = A_d + (-1)^t \Delta A_d, w_{\sigma_t,d} = A_d - (-1)^t \Delta A_d,\\
&h_{\sigma_t,s} = A_s + (-1)^t \Delta A_s, w_{\sigma_t,s} = A_s - (-1)^t \Delta A_s,\\
&h'_{\sigma_t,in} = A_{in} + (-1)^t \Delta A_{in}, w'_{\sigma_t,in} = A_{in} - (-1)^t \Delta A_{in},\\
&h_{\sigma_t,out} = A_{out} + (-1)^t \Delta A_{out}, w_{\sigma_t,out} = A_{out} - (-1)^t \Delta A_{out},\\
&h_{\varphi} = A_{\varphi} + (-1)^t \Delta A_{\varphi}, w_{\varphi} = A_{\varphi} - (-1)^t \Delta A_{\varphi}.
\end{aligned} \tag{C-2}$$

where $(w_{\sigma_t,d}, h_{\sigma_t,d})$ and $(w_{\sigma_t,s}, h_{\sigma_t,s})$ respectively represent the dilation rate and stride in the width and height directions after rotating r_{σ_t}, and $(w'_{\sigma_t,in}, h'_{\sigma_t,in})$, $(w_{\sigma_t,out}, h_{\sigma_t,out})$, and $(h_{\sigma_t,\varphi}, w_{\sigma_t,\varphi})$ respectively represent the size of input feature map (after padding), output feature map, and convolution kernel after rotating r_{σ_t}. Note that, for convenience, we omit σ_t when σ_t=σ_0.

So, $R_{t,out}(m), R_{t,in}(n), R_{t,\phi}(m,n)$ in Eq. (20) can be respectively written as Eqs. (C-3)-(C-5), and Φ can be recast as Eq. (C-6).

$$\begin{aligned}
&R_{t,out}(m)\\
&= \begin{bmatrix} A_{out} - \Delta A_{out} & 1 \end{bmatrix} \left(\begin{bmatrix} \cos\frac{\pi t}{2} & \sin\frac{\pi t}{2} \\ -\sin\frac{\pi t}{2} & \cos\frac{\pi t}{2} \end{bmatrix} \begin{bmatrix} \left\langle m/(A_{out} - (-1)^t \Delta A_{out}) \right\rangle - \frac{A_{out} + (-1)^t \Delta A_{out} - 1}{2} \\ \left\{ m | (A_{out} - (-1)^t \Delta A_{out}) \right\} - \frac{A_{out} - (-1)^t \Delta A_{out} - 1}{2} \end{bmatrix} + \begin{bmatrix} \frac{A_{out} + \Delta A_{out} - 1}{2} \\ \frac{A_{out} - \Delta A_{out} - 1}{2} \end{bmatrix} \right)\\
&= \left((A_{out} - \Delta A_{out}) \cos\tfrac{\pi t}{2} \left\langle m/(A_{out} - \Delta A_{out}) \right\rangle + \cos\tfrac{\pi t}{2} \left\{ m | (A_{out} - \Delta A_{out}) \right\} \right)\\
&+ \left(-\sin\tfrac{\pi t}{2} \left\langle m/(A_{out} + \Delta A_{out}) \right\rangle + (A_{out} - \Delta A_{out}) \sin\tfrac{\pi t}{2} \left\{ m | (A_{out} + \Delta A_{out}) \right\} \right)\\
&+ (A_{out} - \Delta A_{out}) \left(-\cos\tfrac{\pi t}{2} - \sin\tfrac{\pi t}{2} + 1 \right) \tfrac{A_{out} + \Delta A_{out} - 1}{2} + \left(-\cos\tfrac{\pi t}{2} + \sin\tfrac{\pi t}{2} + 1 \right) \tfrac{A_{out} - \Delta A_{out} - 1}{2}.
\end{aligned} \tag{C-3}$$

$$\begin{aligned}
&R_{t,in}(n)\\
&= \begin{bmatrix} A_{in} - \Delta A_{in} & 1 \end{bmatrix} \left(\begin{bmatrix} \cos\frac{\pi t}{2} & \sin\frac{\pi t}{2} \\ -\sin\frac{\pi t}{2} & \cos\frac{\pi t}{2} \end{bmatrix} \begin{bmatrix} \left\langle n/(A_{in} - (-1)^t \Delta A_{in}) \right\rangle - \frac{A_{in} + (-1)^t \Delta A_{in} - 1}{2} \\ \left\{ n | (A_{in} - (-1)^t \Delta A_{in}) \right\} - \frac{A_{in} - (-1)^t \Delta A_{in} - 1}{2} \end{bmatrix} + \begin{bmatrix} \frac{A_{in} + \Delta A_{in} - 1}{2} \\ \frac{A_{in} - \Delta A_{in} - 1}{2} \end{bmatrix} \right)\\
&= \left((A_{in} - \Delta A_{in}) \cos\tfrac{\pi t}{2} \left[\left\langle n/(A_{in} - \Delta A_{in}) \right\rangle \right] + \cos\tfrac{\pi t}{2} \left\{ n | (A_{in} - \Delta A_{in}) \right\} \right)\\
&+ \left(-\sin\tfrac{\pi t}{2} \left[\left\langle n/(A_{in} + \Delta A_{in}) \right\rangle \right] + (A_{in} - \Delta A_{in}) \sin\tfrac{\pi t}{2} \left\{ n | (A_{in} + \Delta A_{in}) \right\} \right)\\
&+ (A_{in} - \Delta A_{in}) \left(-\cos\tfrac{\pi t}{2} - \sin\tfrac{\pi t}{2} + 1 \right) \tfrac{A_{in} + \Delta A_{in} - 1}{2} + \left(-\cos\tfrac{\pi t}{2} + \sin\tfrac{\pi t}{2} + 1 \right) \tfrac{A_{in} - \Delta A_{in} - 1}{2}.
\end{aligned} \tag{C-4}$$

$$\begin{aligned}
&R_{t,\varphi}(m,n)\\
&= \left(\begin{bmatrix} \cos\frac{\pi t}{2} & \sin\frac{\pi t}{2} \\ -\sin\frac{\pi t}{2} & \cos\frac{\pi t}{2} \end{bmatrix} \begin{bmatrix} \frac{1}{A_d + (-1)^t \Delta A_d} \left(\left\langle n/(A_{in} - (-1)^t \Delta A_{in}) \right\rangle - \left\langle m | (A_{out} - (-1)^t \Delta A_{out}) \right\rangle \left(A_s + (-1)^t \Delta A_s \right) \right) \\ \frac{1}{A_d - (-1)^t \Delta A_d} \left(\left\{ n | (A_{in} - (-1)^t \Delta A_{in}) \right\} - \left\{ m | (A_{out} - (-1)^t \Delta A_{out}) \right\} \left(A_s - (-1)^t \Delta A_s \right) \right) \end{bmatrix} \right.\\
&\left. \begin{matrix} - \frac{A_\varphi + (-1)^t \Delta A_\varphi - 1}{2} \\ - \frac{A_\varphi - (-1)^t \Delta A_\varphi - 1}{2} \end{matrix} \Bigg] + \begin{bmatrix} \frac{A_\varphi + \Delta A_\varphi - 1}{2} \\ \frac{A_\varphi - \Delta A_\varphi - 1}{2} \end{bmatrix} \right)\\
&= \begin{bmatrix} \cos\frac{\pi t}{2} \frac{\langle n/w'_{in} \rangle - \langle m | w_{out} \rangle h_s}{h_d} + \sin\frac{\pi t}{2} \frac{\langle n/h'_{in} \rangle - \langle m | h_{out} \rangle h_s}{h_d} - (\cos\frac{\pi t}{2} + \sin\frac{\pi t}{2} - 1) \frac{h_\varphi - 1}{2} \\ -\sin\frac{\pi t}{2} \frac{\langle n/h'_{in} \rangle - \langle m | h_{out} \rangle w_s}{w_d} + \sin\frac{\pi t}{2} \frac{\langle n/w'_{in} \rangle - \langle m | w_{out} \rangle w_s}{w_d} - (\cos\frac{\pi t}{2} - \sin\frac{\pi t}{2} - 1) \frac{w_\varphi - 1}{2} \end{bmatrix}.
\end{aligned} \tag{C-5}$$

$$\Phi(m,n) = \varphi \left(\frac{\left\langle n/(A_{in} - \Delta A_{in}) \right\rangle - \left\langle m/(A_{out} - \Delta A_{out}) \right\rangle (A_s + \Delta A_s)}{(A_d + \Delta A_d)}, \frac{\left\{ n | (A_{in} - \Delta A_{in}) \right\} - \left\{ m | (A_{out} - \Delta A_{out}) \right\} (A_s + \Delta A_s)}{(A_d + \Delta A_d)} \right). \tag{C-6}$$

When $m = R_{t,out}(m), n = R_{t,in}(n)$, we can further get Eqs. (C-7)-(C-10).

$$\begin{aligned}
&\langle R_{t,out}(m)/(A_{out}-\Delta A_{out})\rangle = \Big\langle \cos\tfrac{\pi t}{2}\langle m/(A_{out}-\Delta A_{out})\rangle + \sin\tfrac{\pi t}{2}\{m|(A_{out}+\Delta A_{out})\} - \sin\tfrac{\pi t}{2}\tfrac{\langle m/(A_{out}+\Delta A_{out})\rangle}{A_{out}-\Delta A_{out}} \\
&+\cos\tfrac{\pi t}{2}\tfrac{\{m|(A_{out}-\Delta A_{out})\}}{A_{out}-\Delta A_{out}} + \left(-\cos\tfrac{\pi t}{2}-\sin\tfrac{\pi t}{2}+1\right)\tfrac{A_{out}+\Delta A_{out}-1}{2} + +\tfrac{1}{A_{out}-\Delta A_{out}}\left(\sin\tfrac{\pi t}{2}-\cos\tfrac{\pi t}{2}+1\right)\tfrac{A_{out}-\Delta A_{out}-1}{2}\Big\rangle \\
&= \begin{cases}
\langle m/(A_{out}-\Delta A_{out})\rangle, t=0 \\
\Big\langle \{m|(A_{out}+\Delta A_{out})\} - \frac{\langle m/(A_{out}+\Delta A_{out})\rangle}{A_{out}-\Delta A_{out}} + \frac{A_{out}-\Delta A_{out}-1}{A_{out}-\Delta A_{out}}\Big\rangle = \{m|(A_{out}+\Delta A_{out})\}, t=1 \\
\Big\langle -\frac{m}{A_{out}-\Delta A_{out}} + A_{out}+\Delta A_{out}-1 + \frac{A_{out}-\Delta A_{out}-1}{A_{out}-\Delta A_{out}} - \frac{\{m|(A_{out}-\Delta A_{out})\}}{A_{out}-\Delta A_{out}}\Big\rangle = A_{out}+\Delta A_{out}-1-\Big\langle \frac{m}{A_{out}-\Delta A_{out}}\Big\rangle, t=2 \\
\Big\langle -\{m|(A_{out}+\Delta A_{out})\} + A_{out}+\Delta A_{out}-1 + \frac{\langle m/(A_{out}+\Delta A_{out})\rangle}{A_{out}-\Delta A_{out}}\Big\rangle = A_{out}+\Delta A_{out}-1-\{m|(A_{out}+\Delta A_{out})\}, t=3
\end{cases} \\
&= \cos\tfrac{\pi t}{2}\langle m/(A_{out}-\Delta A_{out})\rangle + \sin\tfrac{\pi t}{2}\{m|(A_{out}+\Delta A_{out})\} - \left(\cos\tfrac{\pi t}{2}+\sin\tfrac{\pi t}{2}-1\right)\tfrac{A_{out}+\Delta A_{out}-1}{2}.
\end{aligned} \tag{C-7}$$

$$\begin{aligned}
\{R_{t,out}(m)|(A_{out}-\Delta A_{out})\} &= \begin{cases}
\{m|(A_{out}-\Delta A_{out})\}, t=0 \\
-\langle m|(A_{out}+\Delta A_{out})\rangle + A_{out}-\Delta A_{out}-1, t=1 \\
-\{m|(A_{out}-\Delta A_{out})\} + A_{out}-\Delta A_{out}-1, t=2 \\
\langle m|(A_{out}+\Delta A_{out})\rangle, t=3
\end{cases} \\
&= \cos\tfrac{\pi t}{2}\langle m/(A_{out}-\Delta A_{out})\rangle + \sin\tfrac{\pi t}{2}\{m|(A_{out}+\Delta A_{out})\} - \left(\cos\tfrac{\pi t}{2}-\sin\tfrac{\pi t}{2}-1\right)\tfrac{A_{out}-\Delta A_{out}-1}{2},
\end{aligned} \tag{C-8}$$

$$\langle R_{t,in}(n)/(A_{in}-\Delta A_{in})\rangle = \cos\tfrac{\pi t}{2}\langle n/(A_{in}-\Delta A_{in})\rangle + \sin\tfrac{\pi t}{2}\{n|(A_{in}+\Delta A_{in})\} - \left(\cos\tfrac{\pi t}{2}+\sin\tfrac{\pi t}{2}-1\right)\tfrac{A_{in}+\Delta A_{in}-1}{2}, \tag{C-9}$$

$$\{R_{t,in}(n)|(A_{in}-\Delta A_{in})\} = \cos\tfrac{\pi t}{2}\langle n|(A_{in}-\Delta A_{in})\rangle + \sin\tfrac{\pi t}{2}\{n|(A_{in}+\Delta A_{in})\} - \left(\cos\tfrac{\pi t}{2}-\sin\tfrac{\pi t}{2}-1\right)\tfrac{A_{in}-\Delta A_{in}-1}{2}. \tag{C-10}$$

Finally, we substitute Eqs. (C-7)-(C-10) into Eq. (C-6), and get Eq. (C-11).

$$\begin{aligned}
&\Phi(R_{t,out}(m), R_{t,in}(n)) = \\
&\varphi\left(\frac{\langle R_{\sigma_t,in}(n)/(A_{in}-\Delta A_{in})\rangle - \langle R_{\sigma_t,out}(m)/(A_{in}-\Delta A_{in})\rangle h_s}{h_d}, \frac{\{R_{\sigma_t,in}(n)/(A_{in}-\Delta A_{in})\} - \{R_{\sigma_t,out}(m)/(A_{in}-\Delta A_{in})\} w_s}{w_d}\right) \\
&= \varphi\Bigg(\cos\tfrac{\pi t}{2}\frac{\langle n/w'_{in}\rangle - \langle m|w_{out}\rangle h_s}{h_d} + \sin\tfrac{\pi t}{2}\frac{\langle n/h'_{in}\rangle - \langle m|h_{out}\rangle h_s}{h_d} - \left(\cos\tfrac{\pi t}{2}+\sin\tfrac{\pi t}{2}-1\right)\left(\frac{(h'_{in}-1)-(h_{out}-1)h_s}{2h_d}\right), \\
&-\sin\tfrac{\pi t}{2}\frac{\langle n/h'_{in}\rangle - \langle m|h_{out}\rangle w_s}{w_d} + \sin\tfrac{\pi t}{2}\frac{\langle n/w'_{in}\rangle - \langle m|w_{out}\rangle w_s}{w_d} - \left(\cos\tfrac{\pi t}{2}-\sin\tfrac{\pi t}{2}-1\right)\left(\frac{(w'_{in}-1)-(w_{out}-1)w_s}{2w_d}\right)\Bigg)
\end{aligned} \tag{C-11}$$

Through comparing Eq. (C-11) and Eq. (C-5), we can find

$$\begin{aligned}
w_{out} &= (w'_{in} - w_d(w_\varphi - 1) - 1)/w_s + 1 = (p_l + p_r + w_{in} - w_d(w_\varphi - 1) - 1)/w_s + 1, \\
h_{out} &= (h'_{in} - h_d(h_\varphi - 1) - 1)/h_s + 1 = (p_a + p_b + h_{in} - h_d(h_\varphi - 1) - 1)/h_s + 1.
\end{aligned} \tag{C-12}$$

The proof of Eq. (21a) is complete. Obviously, in Eq. (21a), researchers can choose different padding according to the practical requirement, which also means Eq. (21a) is not limited by the sizes and hyparameters.

Appendix D

In this appendix, we will deduce Eq. (21b) according to Eq. (20) (i.e., $F_{in}(R_{t,in}(n)) = [F_{in}]_t(n), h_{out}w_{out} = h_{\sigma_t,out}w_{\sigma_t,out}$). First, we respectively calculate $F_{in}(R_{t,in}(n))$ and $[F_{in}]_t(n)$, as shown in Eqs. (D-1)-(D-2).

$$\begin{aligned}
&F_{in}(R_{t,in}(n)) = f_{in}\left(\langle R_{t,in}(n)/w'_{in}\rangle - p_a, \{R_{t,in}(n)|w'_{in}\} - p_l\right) \\
&= f_{in}\left(\langle R_{t,in}(n)/(A_{in}-\Delta A_{in})\rangle - p_a, \{R_{t,in}(n)|(A_{in}-\Delta A_{in})\} - p_l\right) \\
&= f_{in}\begin{pmatrix} \cos\frac{\pi t}{2}\langle n/w'_{in}\rangle + \sin\frac{\pi t}{2}\{n|h'_{in}\} - (\cos\frac{\pi t}{2}+\sin\frac{\pi t}{2}-1)\frac{h'_{in}-1}{2} - p_a, \\ -\sin\frac{\pi t}{2}\langle n/w'_{in}\rangle + \cos\frac{\pi t}{2}\{n|h'_{in}\} - (\cos\frac{\pi t}{2}-\sin\frac{\pi t}{2}-1)\frac{w'_{in}-1}{2} - p_l \end{pmatrix},
\end{aligned} \tag{D-1}$$

$$
\begin{aligned}
&[F_{in}]_t(n)(n)\\
&= flatten((r_{\sigma_t} f_{in}) \oplus (p_{\sigma_t,a}, p_{\sigma_t,b}, p_{\sigma_t,l}, p_{\sigma_t,r}))(n)\\
&= r_{\sigma_t} f_{in}\left(\left\langle n/w'_{\sigma_t,in}\right\rangle - p_{\sigma_t,a}, \left\{n|w'_{\sigma_t,in}\right\} - p_{\sigma_t,l}\right)\\
&= f_{in}\left(\left[\begin{array}{cc} \cos\frac{\pi t}{2} & \sin\frac{\pi t}{2} \\ -\sin\frac{\pi t}{2} & \cos\frac{\pi t}{2} \end{array}\right]\left[\begin{array}{c} \left\langle n/w'_{\sigma_t,in}\right\rangle - p_{\sigma_t,a} - \frac{h_{\sigma_t,in}-1}{2} \\ \left\{n/w'_{\sigma_t,in}\right\} - p_{\sigma_t,l} - \frac{w_{\sigma_t,in}-1}{2} \end{array}\right] + \left[\begin{array}{c} \frac{h_{in}-1}{2} \\ \frac{w_{in}-1}{2} \end{array}\right]\right)\\
&= f_{in}\left(\left[\begin{array}{c} \cos\frac{\pi t}{2}\left\langle n/w'_{in}\right\rangle + \sin\frac{\pi t}{2}\left\{n|h'_{in}\right\} - (\cos\frac{\pi t}{2} + \sin\frac{\pi t}{2} - 1)\frac{h_{in}-1}{2} - \left(\cos\frac{\pi t}{2}p_{\sigma_t,a} + \sin\frac{\pi t}{2}p_{\sigma_t,l}\right) \\ -\sin\frac{\pi t}{2}\left\langle n/w'_{in}\right\rangle + \cos\frac{\pi t}{2}\left\{n|h'_{in}\right\} - (\cos\frac{\pi t}{2} - \sin\frac{\pi t}{2} - 1)\frac{w_{in}-1}{2} - \left(-\sin\frac{\pi t}{2}p_{\sigma_t,a} + \cos\frac{\pi t}{2}p_{\sigma_t,l}\right) \end{array}\right]\right)\\
&= f_{in}\left(\left[\begin{array}{c} \cos\frac{\pi t}{2}\left\langle n/w'_{in}\right\rangle + \sin\frac{\pi t}{2}\left\{n|h'_{in}\right\} - (\cos\frac{\pi t}{2} + \sin\frac{\pi t}{2} - 1)\frac{h'_{in}-1}{2} \\ -\sin\frac{\pi t}{2}\left\langle n/w'_{in}\right\rangle + \cos\frac{\pi t}{2}\left\{n|h'_{in}\right\} - (\cos\frac{\pi t}{2} - \sin\frac{\pi t}{2} - 1)\frac{w'_{in}-1}{2} \end{array}\right.\right.\\
&\quad\left.\left.\begin{array}{c} -\left(\cos\frac{\pi t}{2}p_{\sigma_t,a} + \sin\frac{\pi t}{2}p_{\sigma_t,l} - \frac{1}{2}(\cos\frac{\pi t}{2} + \sin\frac{\pi t}{2} - 1)(p_a + p_b)\right) \\ -\left(-\sin\frac{\pi t}{2}p_{\sigma_t,a} + \cos\frac{\pi t}{2}p_{\sigma_t,l} - \frac{1}{2}(\cos\frac{\pi t}{2} - \sin\frac{\pi t}{2} - 1)(p_l + p_r)\right) \end{array}\right]\right).
\end{aligned}
\tag{D-2}
$$

Then, as $F_{in}(R_{t,in}(n)) = [F_{in}]_t(n)$ (i.e., Eq. (D-1) = Eq. (D-2)), we can easily attain

$$
\begin{aligned}
&\left[\begin{array}{c} p_a \\ p_l \end{array}\right] = \left[\begin{array}{cc} \cos\frac{\pi t}{2} & \sin\frac{\pi t}{2} \\ -\sin\frac{\pi t}{2} & \cos\frac{\pi t}{2} \end{array}\right]\left[\begin{array}{c} p_{\sigma_t,a} \\ p_{\sigma_t,l} \end{array}\right] - \left[\begin{array}{c} \frac{1}{2}(\cos\frac{\pi t}{2} + \sin\frac{\pi t}{2} - 1)(p_a + p_b) \\ \frac{1}{2}(\cos\frac{\pi t}{2} - \sin\frac{\pi t}{2} - 1)(p_l + p_r) \end{array}\right],\\
&\left[\begin{array}{c} p_{\sigma_t,a} \\ p_{\sigma_t,l} \end{array}\right] = \left[\begin{array}{cc} \cos\frac{\pi t}{2} & -\sin\frac{\pi t}{2} \\ \sin\frac{\pi t}{2} & \cos\frac{\pi t}{2} \end{array}\right],\\
&\left[\begin{array}{cccc} \frac{1}{2}(\cos\frac{\pi t}{2} + \sin\frac{\pi t}{2} + 1) & 0 & \frac{1}{2}(\cos\frac{\pi t}{2} + \sin\frac{\pi t}{2} - 1) & 0 \\ 0 & \frac{1}{2}(\cos\frac{\pi t}{2} - \sin\frac{\pi t}{2} + 1) & 0 & \frac{1}{2}(\cos\frac{\pi t}{2} - \sin\frac{\pi t}{2} - 1) \end{array}\right]\left[\begin{array}{c} p_a \\ p_l \\ p_b \\ p_r \end{array}\right].
\end{aligned}
\tag{D-3}
$$

From Eq. (D-3), we can further get

$$
\left[\begin{array}{c} p_{\sigma_t,a} \\ p_{\sigma_t,l} \end{array}\right] = \left[\begin{array}{cccc} \delta(t) & \delta(t-3) & \delta(t-2) & \delta(t-1) \\ \delta(t-1) & \delta(t) & \delta(t-3) & \delta(t-2) \end{array}\right]\left[\begin{array}{c} p_a \\ p_l \\ p_b \\ p_r \end{array}\right].
\tag{D-4}
$$

Since $h_{out}w_{out} = h_{\sigma_t,out}w_{\sigma_t,out}$, it is deduced that

$$
p_{\sigma_t,a} + p_{\sigma_t,b} = \left\{\begin{array}{l} p_a + p_b, t = 0, 2 \\ p_r + p_l, t = 1, 3 \end{array}\right. .
\tag{D-5}
$$

So, the following formula holds.

$$\begin{bmatrix} p_{\sigma_t,a} \\ p_{\sigma_t,l} \\ p_{\sigma_t,b} \\ p_{\sigma_t,r} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \delta(t)+\delta(t-2) & \delta(t-1)+\delta(t-3) & \delta(t)+\delta(t-2) & \delta(t-1)+\delta(t-3) \\ \delta(t-1)+\delta(t-3) & \delta(t)+\delta(t-2) & \delta(t-1)+\delta(t-3) & \delta(t)+\delta(t-2) \end{bmatrix} \begin{bmatrix} p_a \\ p_l \\ p_b \\ p_r \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} p_{\sigma_t,a} \\ p_{\sigma_t,l} \end{bmatrix} \tag{D-6}$$

$$= \begin{bmatrix} \delta(t) & \delta(t-3) & \delta(t-2) & \delta(t-1) \\ \delta(t-1) & \delta(t) & \delta(t-3) & \delta(t-2) \\ \delta(t-2) & \delta(t-1) & \delta(t) & \delta(t-3) \\ \delta(t-3) & \delta(t-2) & \delta(t-1) & \delta(t) \end{bmatrix} \begin{bmatrix} p_a \\ p_l \\ p_b \\ p_r \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}^t \begin{bmatrix} p_a \\ p_l \\ p_b \\ p_r \end{bmatrix}.$$

The proof for Eq. (21b) is complete. According to Eq. (21b), researchers can calculate the corresponding padding form under σ_t when using our proposed rotation equivariant convolution-group framework.

APPENDIX E

In this appendix, we provide a concrete usage example to guide readers on how to correctly apply our theory.

In the methodology section of the article, relying on the padding technique, we conducted a series of rigorous derivations and proved that for images and convolution kernels of any size, the following formula must be followed to achieve rotation equivariance:

$$\begin{aligned} &p_a = p_{ab} - p_b, \\ &p_r = p_{rl} - p_l, \\ &p_b = \langle p_{ab}|2\rangle, \\ &p_l = \langle p_{rl}|2\rangle, \\ &p_{ab} = (h_{out}-1)\, h_s + h_d(h_\varphi - 1) + 1 - h_{in}, \\ &p_{rl} = (w_{out}-1)\, w_s + w_d(w_\varphi - 1) + 1 - w_{in}, \end{aligned} \tag{E-1}$$

where $\langle A\rangle$ means the downward rounded result of A and the symbol $\{A|B\}$ is the remainder of A divided by B, $(w_s, h_s), (w_d, h_d)$ is the stride and dilation in the directions of width and height, respectively, (w_{out}, h_{out}) and (w_{in}, h_{in}) are the width and height of the output and input feature maps, respectively. (w_φ, h_φ) is the width and height of convolution kernel, $P_{\sigma_i} = [p_{\sigma_i,a}, p_{\sigma_i,b}, p_{\sigma_i,l}, p_{\sigma_i,r}]$ represents padding method in σ_i mode, where $p_{\sigma_i,a}$ is the padding above image, $p_{\sigma_i,b}$ is the padding below image, $p_{\sigma_i,l}$ is the padding on the left, and $p_{\sigma_i,r}$ is the padding on the right. Note that, when $\sigma_i = \sigma_0$, the subscript associated with the convolution mode is omitted.

Furthermore, given that our framework aims to extract feature information from different directions, we have further derived the conditions that padding needs to satisfy within the proposed framework:

$$\begin{bmatrix} p_{\sigma_i,a} \\ p_{\sigma_i,l} \\ p_{\sigma_i,b} \\ p_{\sigma_i,r} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}^i \begin{bmatrix} p_a \\ p_l \\ p_b \\ p_r \end{bmatrix}. \tag{E-2}$$

In order to verify the feasibility and validity of the above conclusions, we demonstrate the specific application of the algorithm in a three-layer network, as shown in Fig. E-1. We utilized asymmetric convolutional kernels to downsample images from a resolution of 6×6 to 3×3, and then upsampled them back to their original resolution through transposed convolution.

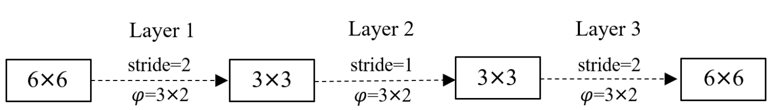

Fig. E-1. The framework of a four-layer network.

The first layer: According to Eq.(E-1), when the image size $w_{in} \times h_{in} = 6 \times 6$, $w_{out} \times h_{out} = 3 \times 3$, convolution kernel size $w_\varphi \times h_\varphi = 3 \times 2$, and dialtion $w_d \times h_d = 1 \times 1$, stride $w_s \times h_s = 2 \times 2$, we can easily obtain the padding in σ_0 mode should be $p_a = p_b = p_l = 0, p_r = 1$. Then, according to Eq.(E-2), we can get different padding methods for the four modes. After passing through the first layer, we can obtain four feature maps from different directions. The details are shown in the Fig. E-2-A:

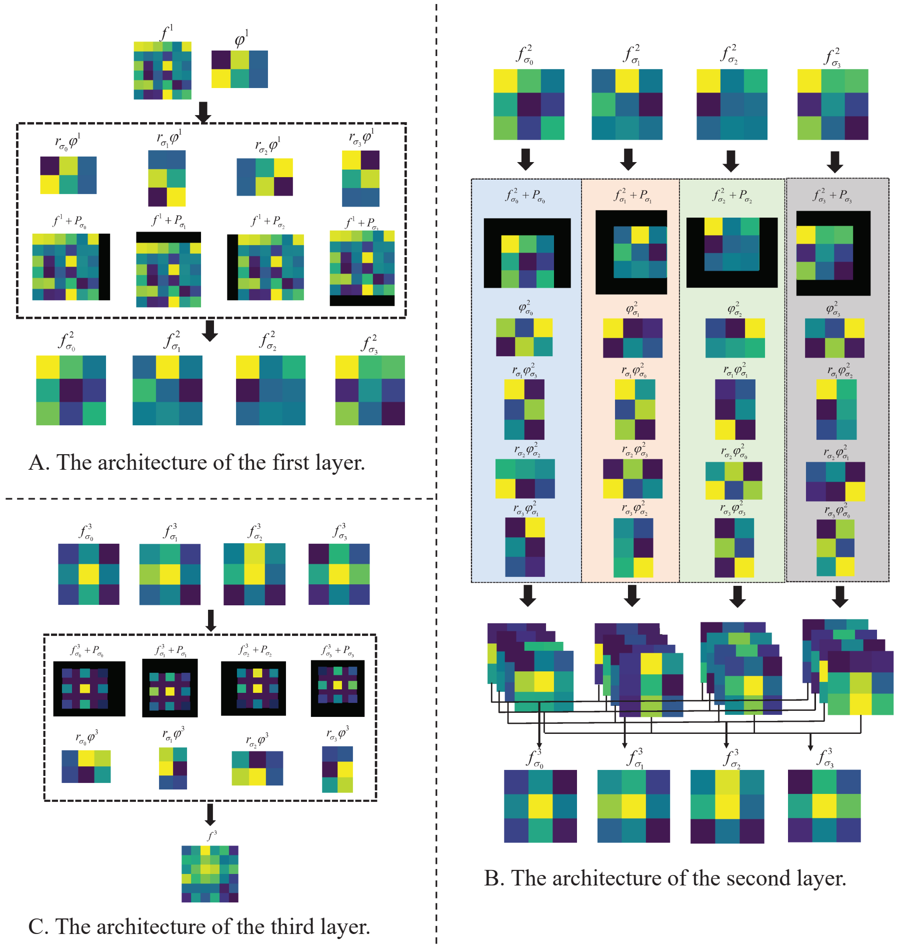

Fig. E-2. The architecture of the first, second, and third layer.

The second layer: Similarly, according to Eq.(E-1), when the image size $w_{in} \times h_{in} = 3 \times 3$, $w_{out} \times h_{out} = 3 \times 3$, convolution kernel size $w_\varphi \times h_\varphi = 3 \times 2$, and dialtion $w_d \times h_d = 1 \times 1$, stride $w_s \times h_s = 2 \times 2$, we can easily obtain the padding in σ_0 mode should be $p_r = p_l = p_a = 1, p_b = 0$. Then, According to Eq.(E-2), We can get different padding methods for the four modes. This layer contains four different convolutional kernels and corresponds to the four feature maps obtained from the first layer. The specific implementation details are shown as Fig. E-2-B.

The third layer: In the final layer, we utilize the transposed convolution to retrieve a resolution of 6×6 from a resolution of 3×3. In this process, the image is first padded to 5×5 when stride = 2, and then apply the same method to calculate the required padding($p_l = p_a = p_b = 1, p_r = 2$). The specific process is shown as Fig. E-2-C.

When a 90° rotated image is input into the network, we observe that the output feature map exhibits a rotation of the same angle, as illustrated in the Fig. E-3. This result strongly validates the effectiveness of our algorithm.

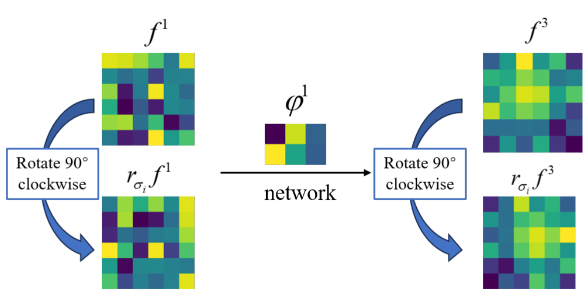

Fig. E-3. The output feature maps of the network before and after the rotation of the input image.