Appendix of
On the Structural Sensitivity of Deep Convolutional Networks to the Directions of Fourier Basis Functions

A. Notations

Notations are summarized in table 1.

B. Preliminary

B.1. Circulant matrix

Let c be a vector and c_i be the i-th element of the vector c. A circulant matrix is a matrix with the following shape.

$$\mathrm{Circ}(c_:) = \begin{bmatrix} c_0 & c_1 & \dots & c_{n-2} & c_{n-1} \\ c_{n-1} & c_0 & c_1 & & c_{n-2} \\ \vdots & c_{n-1} & c_0 & \ddots & \vdots \\ c_2 & & \ddots & \ddots & c_1 \\ c_1 & c_2 & \dots & c_{n-1} & c_0 \end{bmatrix}. \tag{1}$$

A doubly block circulant matrix is a block matrix whose blocks are circulant. The matrix A below is an example of a doubly block circulant matrix.

$$A = \begin{bmatrix} \mathrm{Circ}(K_{0,:}) & \mathrm{Circ}(K_{1,:}) & \dots & \mathrm{Circ}(K_{n-2,:}) & \mathrm{Circ}(K_{n-1,:}) \\ \mathrm{Circ}(K_{n-1,:}) & \mathrm{Circ}(K_{0,:}) & \mathrm{Circ}(K_{1,:}) & & \mathrm{Circ}(K_{n-2,:}) \\ \vdots & \mathrm{Circ}(K_{n-1,:}) & \mathrm{Circ}(K_{0,:}) & \ddots & \vdots \\ \mathrm{Circ}(K_{2,:}) & & \ddots & \ddots & \mathrm{Circ}(K_{1,:}) \\ \mathrm{Circ}(K_{1,:}) & \mathrm{Circ}(K_{2,:}) & \dots & \mathrm{Circ}(K_{n-1,:}) & \mathrm{Circ}(K_{0,:}) \end{bmatrix}, \tag{2}$$

where $K_{i,:}$ is a i-th row of a matrix K. When the channel size of a convolutional layer is equal to one and padding is "wraps around," convolution operation can be written as a doubly block circulant matrix [1, 12].

C. Proof of propositions

C.1. Proposition 1

We prove the proposition following Sedghi *et al.* [12]. Our assumption is that the padding is "wrap around". Under the assumption, a convolutional can be represented by the following matrix M.

$$M = \begin{bmatrix} B^{(0,0)} & B^{(0,1)} & \dots & B^{(0,m_{\mathrm{in}}-1)} \\ B^{(1,0)} & B^{(1,1)} & \dots & B^{(1,m_{\mathrm{in}}-1)} \\ \vdots & \vdots & \ddots & \vdots \\ B^{(m_{\mathrm{out}}-1,0)} & B^{(m_{\mathrm{out}}-1,1)} & \dots & B^{(m_{\mathrm{out}}-1,m_{\mathrm{in}}-1)} \end{bmatrix}, \tag{3}$$

Table 1. Notation table.

$\mathrm{Circ}(c)$: A Circulant matrix crated by a vector c.
x_i: An i-th element of a vector x.
$A_{i,j}$: An i-th row j-th column element of a matrix A.
$A_{i,j}$: An i-th row j-th column element of a matrix A.
ω_N: n-th root of 1, $\exp(2\pi\sqrt{-1}/N)$.
ω_N^i: n-th root of 1 power i.
F_N: A matrix which $(F_N)_{i,j} = \omega_N^{(i+j)}$.
$S(X)$: 2d Fourier transformation of a matrix X.
Q_N: A matrix $F_N \otimes F_N$.
$\otimes$: A Kronecker product.
I_m: m-dimensional identity matrix.
R: $I_m \otimes Q_N$.
m: Channel size.

where each $B^{(c,d)}$ is a doubly circulant matrix. Let $D^{(c,d)} = Q_N^{\mathrm{H}} B^{(c,d)} Q_N$. Since $B_{c,d}$ is a doubly circulant matrix, $D^{(c,d)}$ is a diagonal matrix. Now we can write,

$$\left(I_{\mathrm{out}} \otimes Q_N\right)^{\mathrm{H}} M \left(I_{\mathrm{in}} \otimes Q_N\right) = \begin{bmatrix} D^{(0,0)} & D^{(0,1)} & \dots & D^{(0,m_{\mathrm{in}}-1)} \\ D^{(1,0)} & D^{(1,1)} & \dots & D^{(1,m_{\mathrm{in}}-1)} \\ \vdots & \vdots & \ddots & \vdots \\ D^{(m_{\mathrm{out}}-1,0)} & D^{(m_{\mathrm{out}}-1,1)} & \dots & D^{(m_{\mathrm{out}}-1,m_{\mathrm{in}}-1)} \end{bmatrix}. \tag{4}$$

By multiplying $(I_{m_{\mathrm{out}}} \otimes Q_N)$ from left and $(I_{m_{\mathrm{in}}} \otimes Q_N)^{\mathrm{H}}$ from right, we have

$$M = \left(I_{m_{\mathrm{out}}} \otimes Q_N\right) L \left(I_{m_{\mathrm{in}}} \otimes Q_N\right)^{\mathrm{H}}, \tag{5}$$

where

$$L = \begin{bmatrix} D^{(0,0)} & D^{(0,1)} & \dots & D^{(0,m_{\mathrm{in}}-1)} \\ D^{(1,0)} & D^{(1,1)} & \dots & D^{(1,m_{\mathrm{in}}-1)} \\ \vdots & \vdots & \ddots & \vdots \\ D^{(m_{\mathrm{out}}-1,0)} & D^{(m_{\mathrm{out}}-1,1)} & \dots & D^{(m_{\mathrm{out}}-1,m_{\mathrm{in}}-1)} \end{bmatrix}. \quad \square \tag{6}$$

C.2. Proposition 2

We prove the proposition partially following Sedghi *et al.* [12]. Using prop. 1, $M^{(i)}$ can be decomposed as follows.

$$M^{(i)} = \left(I_{m_{i+1}} \otimes Q_N\right) L^{(i)} \left(I_{m_i} \otimes Q_N\right)^{\mathrm{H}}, \tag{7}$$

where $L^(i)$ is a block matrix such that each block is diagonal. Since

$$\left(I_m \otimes Q_N\right)^{\mathrm{H}} \left(I_m \otimes Q_N\right) = I_{mN^2}, \tag{8}$$

we can write M as

$$M = \left(I_{m_{d+1}} \otimes Q_N\right) \left(\prod_{i=1}^{d} L^{(i)}\right) \left(I_{m_1} \otimes Q_N\right)^{\mathrm{H}}, \tag{9}$$

where d is the number of layers. Let

$$L = \prod_{i=1}^{d} L^{(i)} \tag{10}$$

$$= \begin{bmatrix} D^{(0,0)} & D^{(0,1)} & \dots & D^{(0,m_{\mathrm{in}}-1)} \\ D^{(1,0)} & D^{(1,1)} & \dots & D^{(1,m_{\mathrm{in}}-1)} \\ \vdots & \vdots & \ddots & \vdots \\ D^{(m_{\mathrm{out}}-1,0)} & D^{(m_{\mathrm{out}}-1,1)} & \dots & D^{(m_{\mathrm{out}}-1,m_{\mathrm{in}}-1)} \end{bmatrix}. \tag{11}$$

Since all $L^{(i)}$ are block matrix such that all blocks are diagonal. For any $w \in \{1, \ldots, N^2\}$, let $G^{(w)}$ be a matrix such that

$$G^{(w)}_{i,j} = D^{(i,j)}_{w,w}. \tag{12}$$

Let σ be a singular value of $G^{(w)}$ with a left singular vector x and a right singular vector y. We claim that $y \otimes (Q_N)_{:,w}$ is a right singular vector of M. Let e_w be a standard basis vector. Since $D^{(i,j)}$ is diagonal,

$$L(y \otimes e_w) = \sigma(x \otimes e_w). \tag{13}$$

Thus,

$$\begin{aligned} M\left(y \otimes (Q_N)_{:,w}\right) &= (I_{m_{d+1}} \otimes Q_N) L(y \otimes e_w) && (14)\\ &= \sigma(I_{m_{d+1}} \otimes Q_N)(x \otimes e_w) && (15)\\ &= \sigma(x \otimes e_w). && (16) \end{aligned}$$

Let $\tilde{\sigma}$ be another singular value of $G^{(w)}$ with a left singular vector $\tilde{x}$ and a right singular vector $\tilde{y}$. Then,

$$\begin{aligned} (x \otimes ((Q_N)_{:,w}))^{\mathrm{H}} (\tilde{x} \otimes ((Q_N)_{:,w})) &= (x \otimes ((Q_N)_{:,w}))^{\mathrm{H}} (I_{m_{\mathrm{out}}} \otimes Q_N)(I_{m_{\mathrm{out}}} \otimes Q_N)^{\mathrm{H}} (\tilde{x} \otimes ((Q_N)_{:,w})) && (17)\\ &= (x \otimes e_w)^{\mathrm{H}} (\tilde{x} \otimes e_w) && (18)\\ &= x^{\mathrm{H}} \tilde{x} && (19)\\ &= 0. && (20) \end{aligned}$$

Similarly,

$$(y \otimes ((Q_N)_{:,w}))^{\mathrm{H}} (\tilde{y} \otimes ((Q_N)_{:,w})) = 0. \tag{21}$$

Also,

$$\begin{aligned} (x \otimes ((Q_N)_{:,w}))^{\mathrm{H}} (x \otimes ((Q_N)_{:,w})) &= 1, && (22)\\ (y \otimes ((Q_N)_{:,w}))^{\mathrm{H}} (y \otimes ((Q_N)_{:,w})) &= 1. && (23) \end{aligned}$$

Let $\tilde{\sigma}$ be another singular value of $G^{(\tilde{w})}$ with a left singular vector $\tilde{x}$ and a right singular vector $\tilde{y}$, where $w \neq \tilde{w}$. Then,

$$\begin{aligned} (x \otimes ((Q_N)_{:,w}))^{\mathrm{H}} (\tilde{x} \otimes ((Q_N)_{:,\tilde{w}})) &= (x \otimes ((Q_N)_{:,w}))^{\mathrm{H}} (I_{m_{\mathrm{out}}} \otimes Q_N)(I_{m_{\mathrm{out}}} \otimes Q_N)^{\mathrm{H}} (\tilde{x} \otimes ((Q_N)_{:,\tilde{w}})) && (24)\\ &= (x \otimes e_w)^{\mathrm{H}} (\tilde{x} \otimes e_{\tilde{w}}) && (25)\\ &= 0. && (26) \end{aligned}$$

The last line holds because there are no overwrap in non-zero elements in the two vectors. Similarly,

$$(y \otimes ((Q_N)_{:,w}))^{\mathrm{H}} (\tilde{y} \otimes ((Q_N)_{:,\tilde{w}})) = 0. \tag{27}$$

Thus, using the Kronecker product of singular vectors of $G^{(w)}$ and $(Q_N)_{:,w}$ for all w, we may form a singular value decomposition of M. □

C.3. Proposition 3

Let M be a matrix that represents the convolutional layer. When we have a skip connection, the convolution plus the skip connection can be represented as

$$M + I. \tag{28}$$

Since M is a doubly block circulant matrix, $M + I$ is also a doubly block circulant matrix. Thus, we can apply Prop. 2 with the number of layer $d = 1$. □

C.4. Proposition 4

Normalization layers such as batch-normalization layer at test time or weight-normalization layer can be represented by a multiplication of a diagonal matrix whose elements corresponding to the same channels are equal. Thus, convolutional layers followed by such normalization layers can be represented by Eq.(3). Thus, we can apply Prop. 2. □

C.5. Proposition 5

First, we consider a sampling operation to a tensor G such that we sample elements of inputs whose x, y coordinates are in $\{i, j | i = 0 \pmod s \text{ and } j = 0 \pmod s\}$. For simplicity, we consider a convolution with intput output channel sizes are one. We start from analysis of the output of the operation when its input is $(F_N)_{:,a} \otimes (F_N)_{b,:}$. Since i, j-th element of the output is $((F_N)_{:,a} \otimes (F_N)_{b,:})_{i\times s, j\times s}$ and

$$((F_N)_{:,a} \otimes (F_N)_{b,:})_{i\times s, j\times s} = \left((F_{N/s})_{:,(a\%(N/s))} \otimes (F_{N/s})_{(b\%(N/s)),:}\right)_{i,j}, \tag{29}$$

the output is $(F_{N/s})_{:,(a\%(N/s))} \otimes (F_{N/s})_{(b\%(N/s)),:}$. Thus, when we decompose the input x as

$$x = \sum_{a=0}^{N}\sum_{b=0}^{N} \lambda^{(a,b)} (F_N)_{:,a} \otimes (F_N)_{b,:}, \tag{30}$$

and decompose the output y as

$$y = \sum_{a=0}^{N/s}\sum_{b=0}^{N/s} \tilde{\lambda}^{(a,b)} (F_N)_{:,a} \otimes (F_N)_{b,:}, \tag{31}$$

the following equation holds.

$$\tilde{\lambda}^{(a,b)} = \sum_{l=0}^{s}\sum_{r=0}^{s} \lambda^{(a+lN/s, b+rN/s)}. \tag{32}$$

Thus, the sampling operation can be written as

$$Q_{N/s} S Q_N^{\mathrm{H}}, \tag{33}$$

where

$$S = \begin{bmatrix} I_{(N/s)^2} & I_{(N/s)^2} & \dots & I_{(N/s)^2} \end{bmatrix}. \tag{34}$$

The same holds when the input output channel size is $m \geq 1$, and we can write the operation as

$$(I_m \otimes Q_{N/s}) S' (I_m \otimes Q_N)^{\mathrm{H}}, \tag{35}$$

where S' is a block diagonal matrix such that

$$S' = \begin{bmatrix} S & O & \dots & O \\ O & S & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & S \end{bmatrix}. \tag{36}$$

Let M be a matrix that represents a convolutional layer with stride 1. Since a convolutional layer with stride s can be

represented by a multiplication of M followed by the sampling operation, the convolutional layer can be represented by

$$
\begin{aligned}
&(I_{m_{\mathrm{out}}} \otimes Q_{N/s}) S' (I_{m_{\mathrm{in}}} \otimes Q_N)^{\mathrm{H}} M && (37)\\
=&(I_{m_{\mathrm{out}}} \otimes Q_{N/s}) S' L (I_{m_{\mathrm{in}}} \otimes Q_N)^{\mathrm{H}} && (38)\\
=&(I_{m_{\mathrm{out}}} \otimes Q_{N/s}) \begin{bmatrix} S & O & \dots & O \\ O & S & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & S \end{bmatrix} && (39)\\
&\begin{bmatrix} D^{(0,0)} & D^{(0,1)} & \dots & D^{(0,m_{\mathrm{in}}-1)} \\ D^{(1,0)} & D^{(1,1)} & \dots & D^{(1,m_{\mathrm{in}}-1)} \\ \vdots & \vdots & \ddots & \vdots \\ D^{(m_{\mathrm{out}}-1,0)} & D^{(m_{\mathrm{out}}-1,1)} & \dots & D^{(m_{\mathrm{out}}-1,m_{\mathrm{in}}-1)} \end{bmatrix} (I_{m_{\mathrm{in}}} \otimes Q_N)^{\mathrm{H}} && (40)\\
=&(I_{m_{\mathrm{out}}} \otimes Q_{N/s}) \begin{bmatrix} \tilde{S}^{(0,0)} & \tilde{S}^{(0,1)} & \dots & \tilde{S}^{(0,m_{\mathrm{in}}s-1)} \\ \tilde{S}^{(1,0)} & \tilde{S}^{(1,1)} & \dots & \tilde{S}^{(1,m_{\mathrm{in}}s-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{S}^{(m_{\mathrm{out}}-1,0)} & \tilde{S}^{(m_{\mathrm{out}}-1,1)} & \dots & \tilde{S}^{(m_{\mathrm{out}}-1,m_{\mathrm{in}}s-1)} \end{bmatrix} (I_{m_{\mathrm{in}}} \otimes Q_N)^{\mathrm{H}}, && (41)
\end{aligned}
$$

where $\tilde{S}^{(c,d)}$ is a diagonal matrix. For later use, we define a matrix L as

$$
L = \begin{bmatrix} \tilde{S}^{(0,0)} & \tilde{S}^{(0,1)} & \dots & \tilde{S}^{(0,m_{\mathrm{in}}s-1)} \\ \tilde{S}^{(1,0)} & \tilde{S}^{(1,1)} & \dots & \tilde{S}^{(1,m_{\mathrm{in}}s-1)} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{S}^{(m_{\mathrm{out}}-1,0)} & \tilde{S}^{(m_{\mathrm{out}}-1,1)} & \dots & \tilde{S}^{(m_{\mathrm{out}}-1,m_{\mathrm{in}}s-1)} \end{bmatrix}. \tag{42}
$$

For any $w \in \{1, \dots, (N/s)^2\}$, let $G^{(w)}$ be a matrix such that

$$
G^{(w)}_{i,j} = \tilde{S}^{(i,j)}_{w,w}. \tag{43}
$$

Let σ be a singular value of $G^{(w)}$ with a left singular vector x and a right singular vector y. Let us consider the following unique decomposition:

$$
y = \sum_{i=1}^{s^2} y^{(i)} \otimes \tilde{e}_i, \tag{44}
$$

where $\tilde{e}_i$ is a s-dimensional standard basis vector. We claim that

$$
\sum_{i=1}^{s} y^{(i)} \otimes (Q_N)_{:,w+is}. \tag{45}
$$

is a right singular vector of M.

Let e_w be a $(N/s)^2$-dimensional standard basis vector. Since $\tilde{S}^{(i,j)}$ is diagonal,

$$
L(y \otimes e_w) = \sigma(x \otimes e_w). \tag{46}
$$

Thus,

$$
\begin{aligned}
M \left(\sum_{i=1}^{s} y^{(i)} \otimes (Q_N)_{:,w+is} \right) &= (I_{m_{d+1}} \otimes Q_{N/s}) L (y \otimes e_w) && (47)\\
&= \sigma (I_{m_{d+1}} \otimes Q_{N/s}) (x \otimes e_w) && (48)\\
&= \sigma (x \otimes (Q_{N/s})_{:,w}). && (49)
\end{aligned}
$$

Let $\tilde{\sigma}$ be another singular value of $G^{(w)}$ with a left singular vector $\tilde{x}$ and a right singular vector $\tilde{y}$. Then,

$$\left(\sum_{i=1}^{s} y^{(i)} \otimes (Q_N)_{:,w+is}\right)^{\mathrm{H}} \left(\sum_{i=1}^{s} \tilde{y}^{(i)} \otimes (Q_N)_{:,w+is}\right) \tag{50}$$

$$= \left(\sum_{i=1}^{s} y^{(i)} \otimes (Q_N)_{:,w+is}\right)^{\mathrm{H}} (I_m \otimes Q_N)(I_m \otimes Q_N)^{\mathrm{H}} \left(\sum_{i=1}^{s} \tilde{y}^{(i)} \otimes (Q_N)_{:,w+is}\right) \tag{51}$$

$$= \left(\sum_{i=1}^{s} y^{(i)} \otimes \tilde{e}_i \otimes e_w\right)^{\mathrm{H}} \left(\sum_{i=1}^{s} \tilde{y}^{(i)} \otimes \tilde{e}_i \otimes e_w\right) \tag{52}$$

$$= (y \otimes e_w)^{\mathrm{H}} (\tilde{y} \otimes e_w) \tag{53}$$

$$= y^{\mathrm{H}} \tilde{y} \tag{54}$$

$$= 0. \tag{55}$$

Similarly,

$$\left(x \otimes (Q_{N/s})_{:,w}\right)^{\mathrm{H}} \left(\tilde{x} \otimes (Q_{N/s})_{:,w}\right) = 0 \tag{56}$$

Also,

$$\left(x \otimes (Q_{N/s})_{:,w}\right)^{\mathrm{H}} \left(x \otimes (Q_{N/s})_{:,w}\right) = 1, \tag{57}$$

$$\left(\sum_{i=1}^{s} y^{(i)} \otimes (Q_N)_{:,w+is}\right)^{\mathrm{H}} \left(\sum_{i=1}^{s} y^{(i)} \otimes (Q_N)_{:,w+is}\right) = 1. \tag{58}$$

Let $\tilde{\sigma}$ be another singular value of $G^{(\tilde{w})}$ with a left singular vector $\tilde{x}$ and a right singular vector $\tilde{y}$, where $w \neq \tilde{w}$. Then,

$$\left(\sum_{i=1}^{s} y^{(i)} \otimes (Q_N)_{:,w+is}\right)^{\mathrm{H}} \left(\sum_{i=1}^{s} \tilde{y}^{(i)} \otimes (Q_N)_{:,\tilde{w}+is}\right) \tag{59}$$

$$= \left(\sum_{i=1}^{s} y^{(i)} \otimes (Q_N)_{:,w+is}\right)^{\mathrm{H}} (I_{m_{\mathrm{out}}} \otimes Q_N)(I_{m_{\mathrm{out}}} \otimes Q_N)^{\mathrm{H}} \left(\sum_{i=1}^{s} \tilde{y}^{(i)} \otimes (Q_N)_{:,\tilde{w}+is}\right) \tag{60}$$

$$= (y \otimes e_w)^{\mathrm{H}} (\tilde{y} \otimes e_{\tilde{w}}) \tag{61}$$

$$= 0. \tag{62}$$

The last line holds because there are no overwrap in non-zero elements in the two vectors. Similarly,

$$\left(x \otimes (Q_{N/w})_{:,w}\right)^{\mathrm{H}} \left(\tilde{x} \otimes (Q_{N/s})_{:,\tilde{w}}\right) = 0. \tag{63}$$

Thus, using $x \otimes (Q_{N/w})_{:,w}$ and (44) for all w and singular values, we may form a singular value decomposition of M. □

C.6. Proposition 6

Assume x is a vector such that $S(x)_{u,v} = S(x)^*_{N-u,N-v}$. Let $y = S(x)$. Then,

$$x_{u,v} = S^{-1}(y)_{u,v} \tag{64}$$

$$= \frac{1}{N} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} y_{m,n} \exp(2\pi\sqrt{-1}(um + vn)/N) \tag{65}$$

$$= \frac{1}{N} \sum_{(m,n) \in \{0,\ldots,N-1\} \times \{0,\ldots,N-1\}} y_{m,n} \exp(2\pi\sqrt{-1}(um + vn)/N). \tag{66}$$

When N is odd,

$$\frac{1}{N}\sum_{(m,n)\in\{0,\dots,N-1\}\times\{0,\dots,N-1\}} y_{m,n}\exp(2\pi\sqrt{-1}(um+vn)/N) \tag{67}$$

$$=y_{0,0}+\frac{1}{N}\sum_{(m,n)\in\{0,\dots,N-1\}\times\{0,\dots,N-1\}\setminus\{(0,0)\}}\frac{1}{2}\Big(\tag{68}$$

$$y_{m,n}\exp(2\pi\sqrt{-1}(um+vn)/N)+y_{N-m,N-n}\exp(2\pi\sqrt{-1}(u(N-m)+v(N-n))/N)\Big) \tag{69}$$

$$=y_{0,0}+\frac{1}{N}\sum_{(m,n)\in\{0,\dots,N-1\}\times\{0,\dots,N-1\}\setminus\{(0,0)\}}\left(\mathrm{Re}(y_{m,n})\cos(2\pi(um+vn)/N)-\mathrm{Im}(y_{m,n})\sin(2\pi(um+vn)/N)\right). \tag{70}$$

Since $y_{0,0}=y_{0,0}*$, which means $y_{0,0}$ is real, $x_{u,v}$ is real. When N is even,

$$\frac{1}{N}\sum_{(m,n)\in\{0,\dots,N-1\}\times\{0,\dots,N-1\}} y_{m,n}\exp(2\pi\sqrt{-1}(um+vn)/N) \tag{71}$$

$$=y_{0,0}+y_{N/2,N/2}+\frac{1}{N}\sum_{(m,n)\in\{0,\dots,N-1\}\times\{0,\dots,N-1\}\setminus\{(0,0),(N/2,N/2)\}}\frac{1}{2}\Big(\tag{72}$$

$$y_{m,n}\exp(2\pi\sqrt{-1}(um+vn)/N)+y_{N-m,N-n}\exp(2\pi\sqrt{-1}(u(N-m)+v(N-n))/N)\Big) \tag{73}$$

$$=y_{0,0}+y_{N/2,N/2}+\frac{1}{N}\sum_{(m,n)\in\{0,\dots,N-1\}\times\{0,\dots,N-1\}\setminus\{(0,0)\}}\Big(\tag{74}$$

$$\mathrm{Re}(y_{m,n})\cos(2\pi(um+vn)/N)-\mathrm{Im}(y_{m,n})\sin(2\pi(um+vn)/N)\Big). \tag{75}$$

Since $y_{0,0}=y_{0,0}*$ and $y_{N/2,N/2}=y_{N-N/2,N-N/2}*=y_{N/2,N/2}*$, $x_{u,v}$ is real. Thus, when $S(x)_{u,v}=S(x)^*_{N-u,N-v}$ for all u and v, x is real.

Assume x is a real vector. Then,

$$S(x)^*_{N-u,N-v}=\frac{1}{N}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}y_{m,n}\exp(2\pi\sqrt{-1}((N-u)m+(N-v)n)/N)^* \tag{76}$$

$$=\frac{1}{N}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}y_{m,n}\exp(-2\pi\sqrt{-1}((N-u)m+(N-v)n)/N) \tag{77}$$

$$=\frac{1}{N}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1}y_{m,n}\exp(2\pi\sqrt{-1}(um+vn)/N) \tag{78}$$

$$=S(x)_{u,v}.\quad\square \tag{79}$$

D. Evaluation setups

Datasets: We used MNIST [8], fashion-MNIST [15], SVHN [10], CIFAR10, CIFAR100 [5], and ILSVRC2015 [11] as datasets. For CIFAR10 and CIFAR100, as an data augmentation, we padded four pixels on each side and randomly sampled a 32×32 crop from the padded image or its horizontal flip. 4 pixels are padded on each side We then normalized them with the mean and std of each channel. For training on ILSVRC2015, we augmented data following He *et al*. [3]. For training on ILSVRC2015, we rescaled images with its shorter side randomly sampled in $[256, 480]$ and randomly cropped into 224×244 for scale augmentation [13]. We used per-channel subtraction and standard color augmentation [6]. For other datasets, we scaled inputs into the range from zero to one.

Architectures: We used a multi-layer perceptron (MLP) consisting of 1000–1000 hidden layer with ReLU activation, LeNet [7], WideResNet [16], DenseNet-BC [4], and VGG [13] with batch-normalization for evaluations on datasets except for ILSVRC2015. For ILSVRC2015, we used ResNet50 [3], DenseNet, VGG16, and GoogLeNet [14]. For VGG16 and GoogLeNet, we added a batch-normalization layer after each convolution for faster training.

Training details except for ILSVRC2015: We used Nesterov momentum as an optimizer with momentum 0.9, weight decay 0.0005, and batchsize 128 for the experiments. We trained the MLP and LeNet for 50 epochs with an initial learning rate 0.1 decayed by 0.1 at every 10 epochs. We trained WideResNet as follows. For MNIST, fashion-MNIST, and SVHN, we used width factor $k = 4$, layer 16, and dropout ratio 0.4, and trained for 160 epochs with initial learning ratio 0.01 decayed by 0.1 at epoch 80 and 120. For CIFAR10 and CIFAR100, we used width factor $k = 10$, layer 28, and dropout ratio 0.3, and trained for 200 epochs with initial learning ratio 0.1 decayed by 0.1 at epoch 60, 120, and 160. These are the same configuration for SVHN and CIFAR in Zagoruyko and Komodakis [16]. We trained DenseNet-BC with layer 100, growth rate 12, and dropout ratio 0.2.

Training details on ILSVRC2015: We used SGD with momentum 0.9, weight decay 0.0001, and batchsize 256, and trained for 90 epochs for all architectures. For ResNet50, GoogLeNet, and VGG16, we used the same learning rate scheduling and momentum correction used by Goyal *et al*. [2]. For DenseNet121, we set an initial learning rate to 0.1 and multiplied by 0.1 at epoch 30 and 60, following Huang *et al*. [4].

Metric: We used the fool ratio as a metric, which is the percentage of data that models changed its prediction, following Moosavi-Dezfooli *et al*. [9].

References

[1] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. MIT Press, 2016.

[2] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski, A. Kyrola, A. Tulloch, Y. Jia, and K. He. Accurate, Large Minibatch SGD: Training ImageNet in 1 Hour. *CoRR*, abs/1706.02677, 2017.

[3] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *The IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.

[4] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. Densely Connected Convolutional Networks. In *The IEEE Conference on Computer Vision and Pattern Recognition*, pages 2261–2269, 2017.

[5] A. Krizhevsky. Learning Multiple Layers of Features from Tiny Images. *Computer Science Department, University of Toronto, Technical Report*, 2009.

[6] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Advances in Neural Information Processing Systems 25*, pages 1097–1105, 2012.

[7] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based Learning Applied to Document Recognition. In *Proceedings of the IEEE*, pages 2278–2324, 1998.

[8] Y. LeCun, C. Cortes, and C. J. C. Burges. The MNIST Database of Handwritten Digits. *http://yann.lecun.com/exdb/mnist/*, 1998.

[9] S. Moosavi-Dezfooli, A. Fawzi, O. Fawzi, and P. Frossard. Universal adversarial perturbations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1765–1773, 2017.

[10] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng. Reading Digits in Natural Images with Unsupervised Feature Learning. *Neural Information Processing Systems Workshop*, 2011.

[11] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision (IJCV)*, pages 211–252, 2015.

[12] H. Sedghi, V. Gupta, and P. M. Long. The Singular Values of Convolutional Layers. In *International Conference on Learning Representations*, 2019.

[13] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.

[14] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going Deeper with Convolutions. In *The IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[15] H. Xiao, K. Rasul, and R. Vollgraf. Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms. *CoRR*, abs/1708.07747, 2017.

[16] S. Zagoruyko and N. Komodakis. Wide Residual Networks. In *Proceedings of the British Machine Vision Conference*, pages 87.1–87.12, 2016.