A. Operation Descriptions

A.1. Definitions

A convolution operation is formulated as

$$z = W \otimes x + b, \tag{1}$$

where $x \in \mathbb{R}^{c_i \times h \times w}$ and $z \in \mathbb{R}^{c_o \times h \times w}$ are the input and output tensors, respectively, $W \in \mathbb{R}^{c_o \times k \times k \times c_i}$ and $b \in \mathbb{R}^{c_o}$ are the kernel and the bias of the convolution, and k is the kernel size, which is an odd number, c_i, c_o are the numbers of the input and output channels, and h, w are the input height and width. The element of z at (p, s, t) is:

$$z(p, s, t) = b(p) + \sum_{\ell=0}^{c_i-1} \sum_{\eta=-m}^{m} \sum_{\tau=-m}^{m} W(p, \eta, \tau, \ell) \times x(\ell, s+\eta, t+\tau). \tag{2}$$

where $m = \lfloor k/2 \rfloor$. For convenience, except for the middle two dimensions of the convolution kernel, all indices start from 0. The index range of the two dimensions in the middle of the convolution kernel is $[-m, m]$. If either $s + \eta$ or $t + \tau$ is out of index range, $x(\ell, s + \eta, t + \tau)$ is 0.

I is the identity kernel of the convolution operation with:

$$I(p, s, t, q) = \begin{cases} 1, & p = q \text{ and } s = t = 0, \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

We define the concatenation of the feature vectors in the channel dimension as $z = \begin{bmatrix} x \\ y \end{bmatrix}$, where $x \in \mathbb{R}^{c_1 \times h \times w}$, $y \in \mathbb{R}^{c_2 \times h \times w}$, and $z \in \mathbb{R}^{(c_1+c_2) \times h \times w}$. The element of z at (p, s, t) is:

$$z(p, s, t) = \begin{cases} x(p, s, t), & p < c_1, \\ y(p - c_1, s, t), & \text{otherwise.} \end{cases} \tag{4}$$

The concatenation of the convolution kernels in the horizontal direction (*i.e.*, input-dimension concatenation) is defined as $W = \left[W_{h,1}, W_{h,2}\right]$, where $W_{h,1} \in \mathbb{R}^{c_o \times k \times k \times c_{i1}}$, $W_{h,2} \in \mathbb{R}^{c_o \times k \times k \times c_{i2}}$ and $W \in \mathbb{R}^{c_o \times k \times k \times (c_{i1}+c_{i2})}$. The element of W at (p, s, t, q) is:

$$W(p, s, t, q) = \begin{cases} W_{h,1}(p, s, t, q), & q < c_{i1}, \\ W_{h,2}(p, s, t, q - c_{i1}), & \text{otherwise.} \end{cases} \tag{5}$$

Likewise, the concatenation of the convolution kernels in the vertical direction (*i.e.*, output-dimension concatenation) is defined as $W = \begin{bmatrix} W_{v,1} \\ W_{v,2} \end{bmatrix}$, where $W_{v,1} \in \mathbb{R}^{c_{o1} \times k \times k \times c_i}$, $W_{v,2} \in \mathbb{R}^{c_{o2} \times k \times k \times c_i}$ and $W \in \mathbb{R}^{(c_{o1}+c_{o2}) \times k \times k \times c_i}$. The element of W at (p, t, s, q) is:

$$W(p, t, s, q) = \begin{cases} W_{v,1}(p, t, s, q), & p < c_{o1}, \\ W_{v,2}(p - c_{o1}, t, s, q), & \text{otherwise.} \end{cases} \tag{6}$$

The multiplication of the convolution kernels is defined as: $W_3 = W_1 \times W_2$, where $W_1 \in \mathbb{R}^{c_1 \times k_1 \times k_2 \times c_2}$, $W_2 \in \mathbb{R}^{c_2 \times k_3 \times k_4 \times c_3}$, $W_3 \in \mathbb{R}^{c_1 \times (k_1+k_3-1) \times (k_2+k_4-1) \times c_3}$, and k_1, k_2, k_3 and k_4 are all odd numbers. The element of W_3 at (p, s, t, q) is:

$$W_3(p, s, t, q) = \sum_{\ell=0}^{c_2-1} \sum_{\eta=-m_1}^{m_1} \sum_{\tau=-m_2}^{m_2} W_1(p, \eta, \tau, \ell) \times W_2(\ell, s-\eta, t-\tau, q), \tag{7}$$

where $m_i = \lfloor \frac{k_i}{2} \rfloor$ for $i = 1, 2$. The multiplication of the convolution kernel and bias is defined as $b_2 = W \times b_1$, where $W \in \mathbb{R}^{c_o \times k_1 \times k_2 \times c_i}$, $b_1 \in \mathbb{R}^{c_i}$ and $b_2 \in \mathbb{R}^{c_o}$. The element of b_2 at (p) is:

$$b_2(p) = \sum_{\ell=0}^{c_i-1} \sum_{\eta=-m_1}^{m_1} \sum_{\tau=-m_2}^{m_2} W(p, \eta, \tau, \ell) \times b_1(\ell). \tag{8}$$

The convolution kernel multiplication satisfies:

$$W_1 \otimes (W_2 \otimes x + b_2) + b_1 = (W_1 \times W_2) \otimes x + ((W_1 \times b_2) + b_1), \tag{9}$$

which can be proved with the help of Eq. (2). The transformation process of the reparameterization technique can be described by Eq. (9), and the corresponding codes are provided in the supplementary material.

Moreover, the block multiplication of the convolution kernels has the same properties as the block multiplication of the matrices.

A.2. Kernel Properties

For the kernel properties in the Method section, we have the following properties:

- **Property 1** *For any* $x \in \mathbb{R}^{c\times h\times w}$ *and* $I \in \mathbb{R}^{c\times k\times w\times c}$*, we have* $I \otimes x = x$*,*
- **Property 2** *For any* $W_1 \in \mathbb{R}^{c_o\times k\times k\times c_{i1}}$, $W_2 \in \mathbb{R}^{c_o\times k\times k\times c_{i2}}$, $x \in \mathbb{R}^{c_{i1}\times h\times w}$ *and* $y \in \mathbb{R}^{c_{i2}\times h\times w}$*, we have* $[W_1, W_2] \otimes \begin{bmatrix} x \\ y \end{bmatrix} = W_1 \otimes x + W_2 \otimes y$*,*
- **Property 3** *For any* $W_1 \in \mathbb{R}^{c_{o1}\times k\times k\times c_i}$, $W_2 \in \mathbb{R}^{c_{o2}\times k\times k\times c_i}$, $x \in \mathbb{R}^{c_i\times h\times w}$*, we have* $\begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \otimes x = \begin{bmatrix} W_1 \otimes x \\ W_2 \otimes x \end{bmatrix}$*,*
- **Property 4** *For any* $W \in \mathbb{R}^{c_o\times k\times k\times c_i}$*, we have* $I \times W = W$*.*

In the following, we give proofs of the correctness of these four properties.

Correctness of Property 1. Let $z = I \otimes x$, $m = \lfloor k/2 \rfloor$, then the element of z at (p, s, t) is

$$\begin{aligned} z(p,s,t) &= \sum_{\ell=0}^{c-1} \sum_{\eta=-m}^{m} \sum_{\tau=-m}^{m} I(p,\eta,\tau,\ell) \times x(\ell, s+\eta, t+\tau) \\ &= \sum_{\ell=0}^{c-1} \sum_{k_1=-m}^{m} \sum_{k_2=-m}^{m} [\ell = p] \times [\eta = \tau = 0] \times x(\ell, s+\eta, t+\tau) \\ &= x(p,s,t) \end{aligned}$$

where $[expr]$ is 1 when *expr* is true, and 0 otherwise. □

Correctness of Property 2. Let $W = [W_1, W_2]$, $z = \begin{bmatrix} x \\ y \end{bmatrix}$, $\hat{x} = W_1 \otimes x$, $\hat{y} = W_2 \otimes y$ and $\hat{z} = W \otimes z$, where $W \in \mathbb{R}^{c_o\times k\times k\times (c_{i1}+c_{i2})}$, $z \in \mathbb{R}^{(c_{i1}+c_{i2})\times h\times w}$, $\hat{x} \in \mathbb{R}^{c_o\times h\times w}$, $\hat{y} \in \mathbb{R}^{c_o\times h\times w}$, $\hat{z} \in \mathbb{R}^{c_o\times h\times w}$ and $m = \lfloor k/2 \rfloor$. Then the element of $\hat{z}$ at (p, s, t) is:

$$\begin{aligned} \hat{z}(p,s,t) &= \sum_{\ell=0}^{c_{i1}+c_{i2}-1} \sum_{\eta=-m}^{m} \sum_{\tau=-m}^{m} W(p,\eta,\tau,\ell) \times z(\ell, s+\eta, t+\tau) \\ &= \sum_{\ell_1=0}^{c_{i1}-1} \sum_{\eta=-m}^{m} \sum_{\tau=-m}^{m} W_1(p,\eta,\tau,\ell_1) \times x(\ell_1, s+\eta, t+\tau) \\ &\quad + \sum_{\ell_2=0}^{c_{i2}-1} \sum_{\eta=-m}^{m} \sum_{\tau=-m}^{m} W_2(p,\eta,\tau,\ell_2) \times y(\ell_2, s+\eta, t+\tau) \\ &= \hat{x}(p,s,t) + \hat{y}(p,s,t) \end{aligned}$$

□

Correctness of Property 3. Let $W = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$, $z_1 = W_1 \otimes x$, $z_2 = W_2 \otimes x$, $z = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix} \otimes x$, where $W \in \mathbb{R}^{(c_{o1}+c_{o2})\times k\times k\times c_i}$, $z_1 \in \mathbb{R}^{c_{o1}\times h\times w}$, $z_2 \in \mathbb{R}^{c_{o2}\times h\times w}$, $z \in \mathbb{R}^{(c_{o1}+c_{o2})\times h\times w}$ and $m = \lfloor k/2 \rfloor$.

When $0 \leq p < c_{o1}$, the element of z at (p, s, t) is:

$$\begin{aligned} z(p,s,t) &= \sum_{\ell=0}^{c_i-1} \sum_{\eta=-m}^{m} \sum_{\tau=-m}^{m} W(p,\eta,\tau,\ell) \times x(\ell, s+\eta, t+\tau) \\ &= \sum_{\ell=0}^{c_i-1} \sum_{\eta=-m}^{m} \sum_{\tau=-m}^{m} W_1(p,\eta,\tau,\ell) \times x(\ell, s+\eta, t+\tau) \\ &= z_1(p,s,t) \end{aligned}$$

When $c_{o1} \leq p < c_{o1} + c_{o2}$, the element of z at (p, s, t) is:

$$\begin{aligned} z(p,s,t) &= \sum_{\ell=0}^{c_i-1} \sum_{\eta=-m}^{m} \sum_{\tau=-m}^{m} W(p,\eta,\tau,\ell) \times x(\ell, s+\eta, t+\tau) \\ &= \sum_{\ell=0}^{c_i-1} \sum_{\eta=-m}^{m} \sum_{\tau=-m}^{m} W_2(p-c_{o1},\eta,\tau,\ell) \times x(\ell, s+\eta, t+\tau) \\ &= z_2(p-c_{o1}, s, t) \end{aligned}$$

□

Correctness of Property 4. Let $\hat{W} = I \times W$, $m_i = \lfloor \frac{k_i}{2} \rfloor$ for $i = 1, 2$, then the element of $\hat{W}$ at (p, s, t, q) is:

$$\begin{aligned} \hat{W}(p,s,t,q) &= \sum_{\ell=0}^{c_o-1} \sum_{\eta=-m_1}^{m_1} \sum_{\tau=-m_2}^{m_2} I(p,\eta,\tau,\ell) \times W(\ell, s-\eta, t-\tau, q) \\ &= \sum_{\ell=0}^{c_o-1} \sum_{\eta=-m_1}^{m_1} \sum_{\tau=-m_2}^{m_2} ([\eta=\tau=0] \times [\ell=p]) \times W(\ell, s-\eta, t-\tau, q) \\ &= W(p,s,t,q) \end{aligned}$$

□

Remark 1 *The kernels in step 4 of main paper are* $[I, I]$ *and* $\mathrm{diag}(W, \mathit{repeat}(I, n))$*, and the biases are* 0 *and* $\begin{bmatrix} b \\ 0 \end{bmatrix}$ *respectively. Using Eq. (9) and Property 4, the merging process is as follows:*

$$\begin{aligned} & [I, I] \otimes \left(\begin{bmatrix} W & O \\ O & \mathit{repeat}(I, n) \end{bmatrix} \otimes x + \begin{bmatrix} b \\ 0 \end{bmatrix} \right) + 0 \\ = & \left([I, I] \times \begin{bmatrix} W & O \\ O & \mathit{repeat}(I, n) \end{bmatrix} \right) \otimes x + \left([I, I] \times \begin{bmatrix} b \\ 0 \end{bmatrix} + 0 \right) \\ = & [W, \mathit{repeat}(I, n)] \otimes x + b \end{aligned}$$

so the two convolutions are merged into one convolution, whose kernel is $[W, \mathit{repeat}(I, n)]$ *and bias is* b*.*

B. More Qualitative Comparison

This section provides more qualitative comparisons. Different models in the same column have similar inference latencies.

Figure 1. The window edges of images generated by ETDS-L are sharper. Zooming for details.

Figure 2. In the images generated by ETDS-L, the stripes on the wall are much clearer. Zooming for details.

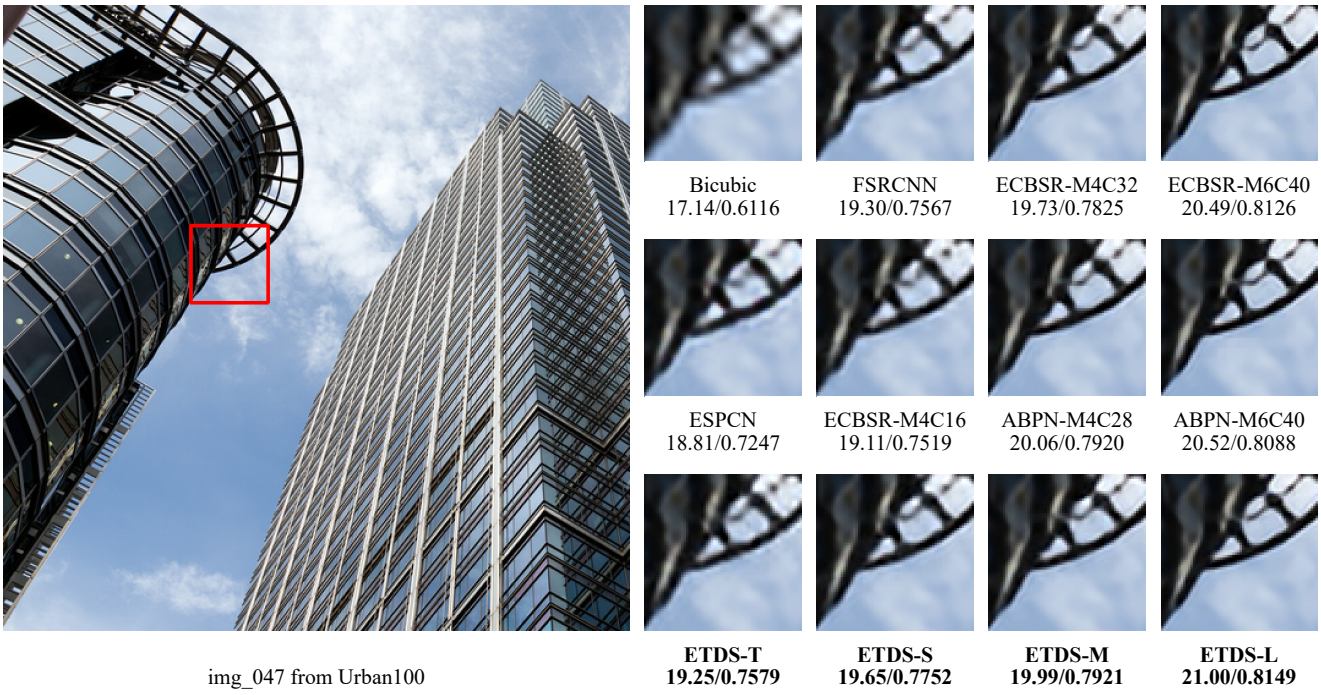

Figure 3. The images generated by ETDS-L have the fewest artifacts. Zooming for details.

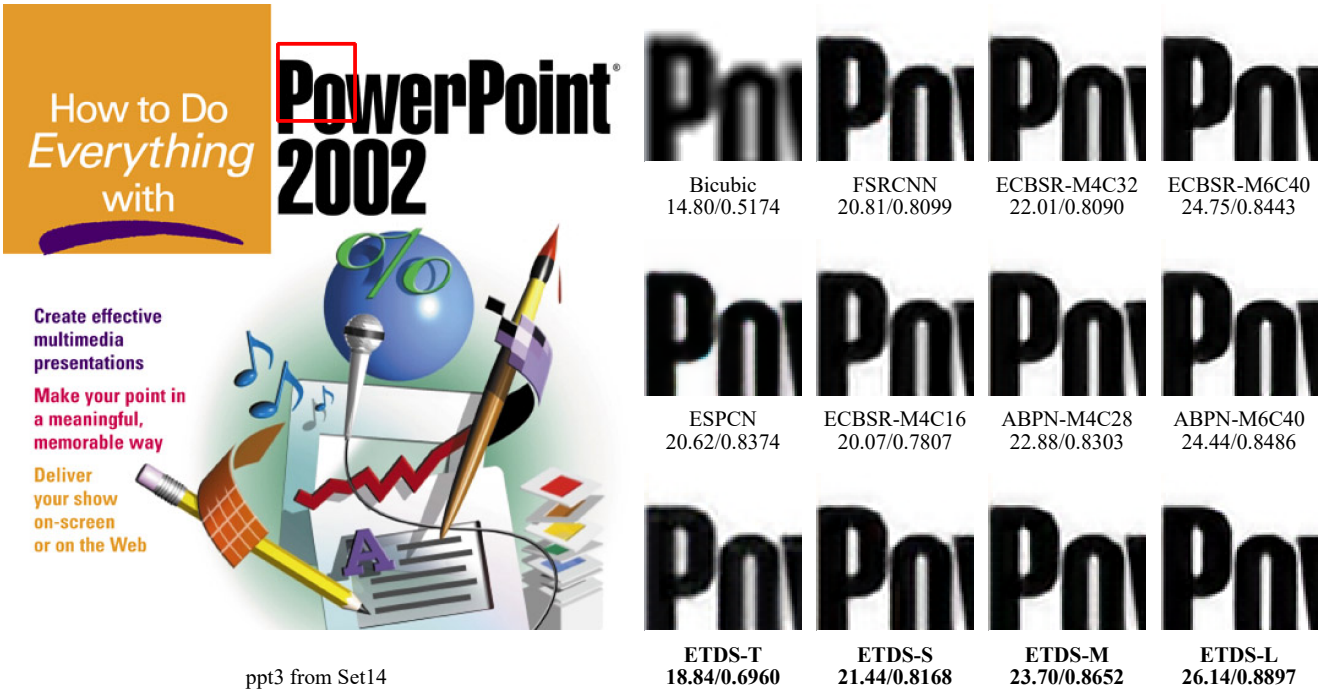

Figure 4. In the images generated by ETDS-L, the edges of the letters are sharper and the area between letter P and letter O has fewer artifacts. Zooming for details.

C. Supplementary Comparison of Ablation Experiments for Equivalent Transformation

Table 1. Statistics of model parameters, MACs, and latency of ECBSR and ABPN with and without ET on $\times 2$, $\times 3$ and $\times 4$ tasks. Better results are marked in **bold**.

Scale	Model	Params (K)	MACs (G)	Latency (ms)		
				CPU	Android NNAPI	MediaTek Neuron
$\times 2$	ECBSR-M4C16	11.55	2.64	208	6.15	4.57
	ECBSR+ET-M4C16	17.47	4.00	**197**	**5.51**	**3.31**
	ECBSR-M4C32	41.52	9.53	**344**	9.70	6.64
	ECBSR+ET-M4C32	52.04	11.94	354	**8.67**	**5.54**
	ECBSR-M6C40	92.37	21.22	617	18.1	10.9
	ECBSR+ET-M6C40	111.27	25.56	**597**	**17**	**9.96**
	ABPN-M4C28	33.46	7.67	185	10.0	6.66
	ABPN+ET-M4C28	44.41	10.19	**182**	**8.6**	**5.27**
	ABPN-M6C40	93.40	21.45	356	18.7	11.2
	ABPN+ET-M6C40	114.29	26.25	**339**	**17.2**	**9.89**
$\times 3$	ECBSR-M4C16	13.72	3.14	225	16.2	11.0
	ECBSR+ET-M4C16	21.02	4.81	**215**	**9.35**	**6.14**
	ECBSR-M4C32	45.85	10.52	356	19.7	13.1
	ECBSR+ET-M4C32	57.90	13.29	**354**	**13.1**	**8.34**
	ECBSR-M6C40	97.79	22.46	652	28.1	17.4
	ECBSR+ET-M6C40	118.28	27.17	**621**	**21.5**	**12.9**
	ABPN-M4C28	42.54	9.76	233	20.5	13.8
	ABPN+ET-M4C28	54.96	12.61	**188**	**13.7**	**8.57**
	ABPN-M6C40	104.10	23.91	378	29.2	18.2
	ABPN+ET-M6C40	126.35	29.02	**373**	**22.4**	**13.5**
$\times 4$	ECBSR-M4C16	16.77	3.83	277	23.9	13.7
	ECBSR+ET-M4C16	26.18	5.99	**202**	**11.7**	**7.79**
	ECBSR-M4C32	51.92	11.91	381	28.0	16.3
	ECBSR+ET-M4C32	65.94	15.13	**360**	**15.7**	**10.3**
	ECBSR-M6C40	105.37	24.20	663	36.8	20.7
	ECBSR+ET-M6C40	127.76	29.34	**627**	**23.5**	**14.5**
	ABPN-M4C28	62.05	14.24	288	30.3	17.4
	ABPN+ET-M4C28	77.78	17.85	**263**	**17.8**	**11.4**
	ABPN-M6C40	125.87	28.91	429	39.1	22.4
	ABPN+ET-M6C40	151.55	34.81	**392**	**26.6**	**16.1**

Table 2. Statistics of model parameters, MACs, and inference latency of our ETDS with and without ET on $\times 2$, $\times 3$ and $\times 4$ tasks. Inference latency is tested on Dimensity 8100C SoC. Better results are marked in **bold**.

Scale	Model	Params (K)	MACs (G)	Latency (ms)					
				CPU		Android NNAPI		MediaTek Neuron	
				INT8	FP32	INT8	FP32	INT8	FP32
$\times 2$	ETDS-T (w/o ET)	12.11	2.75	268	460	10.2	176	6.87	174
	ETDS-T	13.94	3.19	**106**	**187**	**4.85**	**40.1**	**3.81**	**38.1**
	ETDS-S (w/o ET)	38.35	8.77	329	601	13.7	247	9.35	245
	ETDS-S	41.51	9.52	**112**	**391**	**8.06**	**62.2**	**5.77**	**61.0**
	ETDS-M (w/o ET)	55.34	12.65	438	816	16.9	342	12.2	340
	ETDS-M	60.01	13.77	**161**	**511**	**10.2**	**83.6**	**6.73**	**82.5**
	ETDS-L (w/o ET)	143.85	32.98	673	1394	30.3	599	18.4	596
	ETDS-L	152.18	34.97	**342**	**1003**	**20.6**	**175**	**12.1**	**175**
$\times 3$	ETDS-T (w/o ET)	14.30	3.25	298	482	19.6	210	14.2	203
	ETDS-T	16.92	3.86	**97.9**	**204**	**8.64**	**63.8**	**5.36**	**57.4**
	ETDS-S (w/o ET)	42.70	9.76	351	626	22.8	290	16.6	284
	ETDS-S	46.79	10.73	**114**	**390**	**12.2**	**86.8**	**7.73**	**81.3**
	ETDS-M (w/o ET)	59.69	13.65	462	841	26.3	385	19.3	379
	ETDS-M	65.29	14.98	**164**	**514**	**14.4**	**86.5**	**8.76**	**83**
	ETDS-L (w/o ET)	150.36	34.47	711	1441	39.6	642	25.9	635
	ETDS-L	159.77	36.71	**377**	**989**	**25.4**	**202**	**14.2**	**195**
$\times 4$	ETDS-T (w/o ET)	17.37	3.94	333	508	19.2	253	12.4	243
	ETDS-T	21.36	4.88	**116**	**235**	**10.9**	**67.5**	**6.94**	**59.1**
	ETDS-S (w/o ET)	48.79	11.16	372	665	22.8	324	16.0	314
	ETDS-S	54.11	12.41	**125**	**413**	**14.6**	**90.7**	**9.02**	**83.2**
	ETDS-M (w/o ET)	65.78	15.04	490	882	26.7	419	17.9	408
	ETDS-M	72.61	16.66	**175**	**558**	**16.6**	**112**	**9.98**	**104**
	ETDS-L (w/o ET)	159.47	36.56	737	1487	39.9	677	24.0	667
	ETDS-L	169.97	39.05	**373**	**1069**	**27.6**	**205**	**15.3**	**198**

Tab. 1 and Tab. 2 show that once ABPN, ECBSR and ETDS are equipped with ET, the inference latency decreases. Also, it is observed that the inference latency is lower when the scale factor is larger. The possible reason is that with the increase of scale factor, the amount of data transferred by the global residual connection increases, which enlarges the burden on the IO bandwidth and RAM. To this end, tasks of larger scale factors enjoy more benefits of ET.

D. More details about channel mask experiments

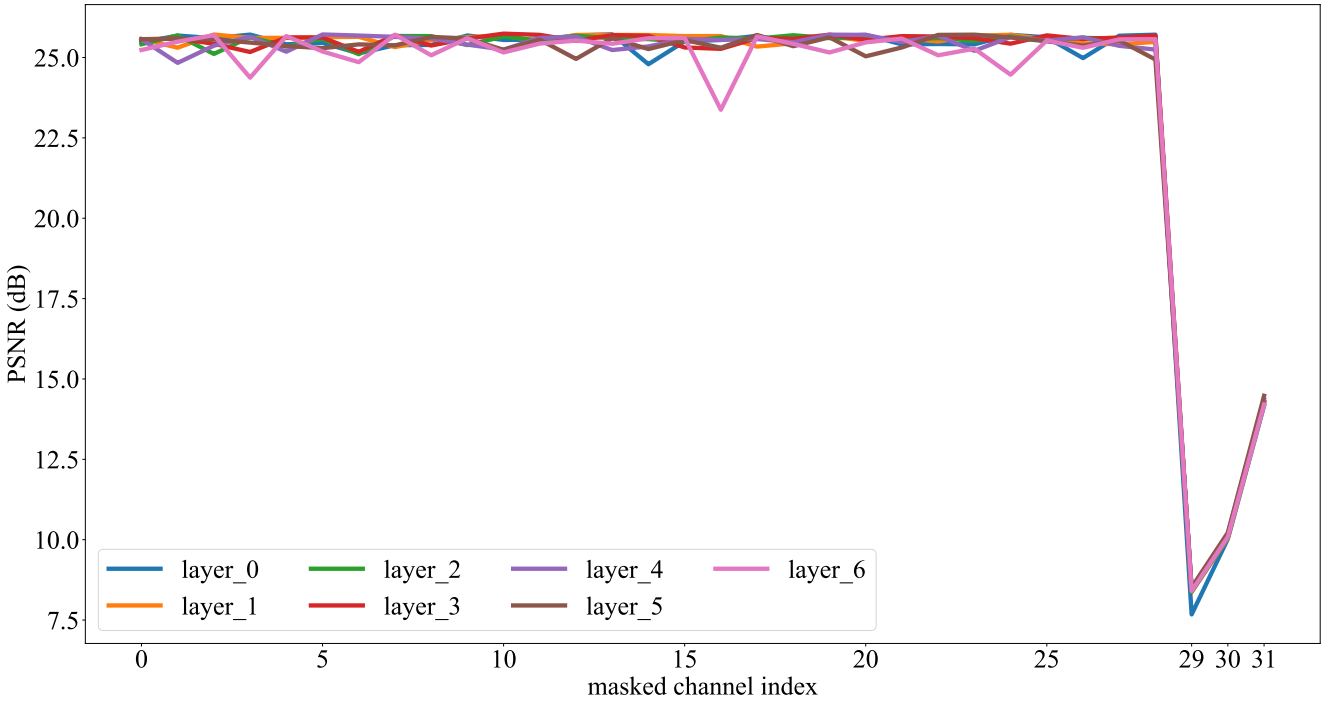

Figure 5. Channel masking experiments for all layers of ETDS.

The high-frequency component of an image is the one where the intensity values change quickly from one pixel to the adjacent ones. On the other hand, the low-frequency counterpart is relatively uniform in brightness or where intensity varies very slowly. Low-frequency components provide the basic information of an image, whose absence causes a significant decrease in image quality. Fig. 5 shows that there is a severe decrease when channels 29, 30 and 31 are masked, which means the low-frequency components are extracted by the last three channels (*i.e.*, residual branch) of each layer, which is consistent with our design goals.